Joint Reconstruction for Multi-Modality Imaging with Common Structure

by

Matthias Joachim Ehrhardt

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Department of Medical Physics and Biomedical Engineering
University College London
April 30, 2015

I, Matthias Joachim Ehrhardt, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Imaging is a powerful tool being used in many disciplines such as engineering, physics, biology and medicine to name a few. Recent years have seen a trend that imaging modalities have been combined to create multi-modality imaging tools where different modalities acquire complementary information. For example, in medical imaging, positron emission tomography (PET) and magnetic resonance imaging (MRI) are combined to image structure and function of the human body. Another example is spectral imaging where each channel provides information about a different wave length, e.g. information about red, green and blue (RGB).

Most imaging modalities do not acquire images directly but measure a quantity from which we can reconstruct an image. These inverse problems require a priori information in order to give meaningful solutions. Assumptions are often on the smoothness of the solution but other information is sometimes available, too. Many multi-modality images show a strong inter-channel correlation as they are acquired from the same anatomy in medical imaging or the same scenery in spectral imaging. However, images from different modalities are usually reconstructed separately.

In this thesis we aim to exploit this correlation using the data from all modalities, that are present in the acquisition, in a joint reconstruction process with the assumption that similar structures in all channels are more likely. We propose a framework for joint reconstruction where modalities are coupled by additional information about the solution we seek. A family of priors – called parallel level sets – allows us to incorporate structural a priori knowledge into the reconstruction. We analyse the parallel level set priors in several aspects including their convexity and the diffusive flow generated by their variation. Several numerical examples in RGB colour imaging and in PET-MRI illustrate the gain of joint reconstruction and in particular of the parallel level set priors.

Acknowledgements

Three intense years come to an end and it is time to close this chapter. As much as I am sad to complete this part of my life, as much am I looking forward to explore the next one - not least due to all the great people who are supporting me.

I am deeply thankful to Prof Simon Arridge for his supervision, inspiration and universal support he gave me during this time. His help, ideas and motivation kept me on track, whatever problem I was facing. Not only am I grateful for your guidance, but I also strongly appreciate that you opened so many doors to me. Thank you Simon, for all your effort.

Dr Kris Thielemans, for the great supervision and expertise he gave me during my PhD. He provided me with all the nasty details about PET a mathematician usually does not want to know, but are so crucial nevertheless. I am very grateful for all your painstaking comments over the years that so much improved everything I wrote, not least this thesis. Thank you for that.

I wish to thank the whole PET-MRI team at UCL for their weekly support. In particular, I want to thank Prof Brian Hutton for his often critical but very helpful comments.

Pawel Markiewicz, Richard Manber, Maria Liljeroth, David Atkinson and Anna Barnes, for all their time and efforts with acquiring and handling data. They never hesitated when we had to do an experiment again and again and again. I learned a great deal due to all their help.

I also greatly acknowledge the funding of my PhD studentship by Siemens and the UCL Faculty of Engineering Sciences. The financial support over the years made all of this possible and enabled me attend and present at conferences all over the globe.

Stefano Pedemonte, for long, long discussions at whatever end of the world and about whatever we just had in mind.

My office mates Luis, Elwin, Felix and Jose. It has been a great pleasure to enjoy the coffee brakes with you – even if I did not win.

I am very thankful to Prof Peter Maass, Dr Stefan Schiffler and Dr Dennis Trede

who guided me during the very early stages of my career. They gave me the motivation, inspiration and opportunity to find my way into inverse problems and image processing.

I also wish to thank my family – my mum Marianne, my dad Helmut with his wife Ute, as well as my brother Guido with his family Melanie and Marlene. I could not have done any of this without the love and support you gave me over the years. Thank you for that.

Last but not least, I wish to thank my dear, honest, lovely, smart and beautiful partner Beate. I am not only very grateful for your time, patience and moral support, but so am I for your trust and faith in me. Actually, I am most thankful for you being always there for me – whatever comes. I love you.

London in April, 2015 Matthias Joachim Ehrhardt

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. Multi-Modality Imaging.

Imaging, the science to create or collect images of an object under investigation, is a powerful tool being used in many disciplines like engineering, physics, biology and medicine to name a few. Recent years have seen a trend that imaging modalities have been combined to create multi-modality imaging tools where different modalities acquire complementary information. There is confidence that "one plus one can add up to more than two" [39].

In medical imaging, for many decades single modality scanners that image either anatomy or function have been available and are nowadays widely used in clinical practice. Anatomy can be imaged for instance by magnetic resonance imaging (MRI) or computer assisted tomography (CT). Often function is measured by radioactive labelled markers using positron emission tomography (PET) or single photon emission computed tomography (SPECT). Based on these single modality scanners, multi-modality scanners were developed to combine the strength of both regimes and image both structure and function. While PET-CT scanners acquire the data for the two modalities sequentially, some recently developed PET-MRI scanners are able to perform both scans *simultaneously* [39, 145, 32, 129, 161, 128, 84]. Moreover, there are attempts to bring this one step further and combine even more than two modalities, so called "omni-tomography" [154].

Another "multi-modality" example is spectral imaging where multiple sensors measure light at different wave lengths. In the simplest case of colour photography each channel contains information about red, green and blue (RGB) [43, 160, 135, 139, 91, 146, 73, 22, 67, 50, 112, 80]. While cameras that capture several channels became available half a century ago, algorithms to reconstruct or analyse colour images became only practical in the 90's mainly due to hardware limitations. Still, to this date

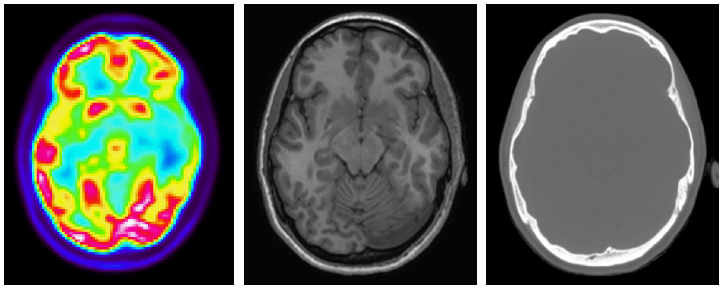


Figure 1.1. *Reconstructions of data acquired by PET, MRI and CT of the same patient. The images were affinely registered to a common frame as they were acquired on a sequential PET-CT and on a separate MRI scanner. Although the images contain completely different biological information, common shapes can be observed due to the same anatomy. Images courtesy of UCL hospital. Figure reproduced from [51].*

many algorithms have been proposed for grey scale images and their only extension to multi-valued images is channel-by-channel.

There are more examples of multi-modality imaging or where multi-valued images are available. These are for instance in medicine: multi-contrast MRI [3, 13], spectral CT [133], biochemistry: imaging mass spectrometry with matrix-assisted laser desorption/ionization (MALDI) [10], and geophysics: seismic ray tomography/seismic travel time inversion and 2D direct current (DC) resistivity [63, 64, 65, 74].

1.2. Inverse Problems and A Priori Information.

Many imaging modalities actually do not acquire images. In medical imaging one usually does not measure directly the quantity of interest as this is often hidden within the body of the patient, e.g. "tissue density" in X-ray CT, the distribution of a radioactive tracer in PET or SPECT or proton density in MRI. In all of these cases the quantity of interest can be linked to the data via a model that depends on the physics and the scanner itself. In many practical cases these models are chosen to be linear. The reconstruction task is then posed as the inversion of such a mapping given noisy measurements. Hadamard defined an *inverse problem* to be well-posed if a solution i) exists, ii) is unique and iii) depends continuously on the data [54]. If one of these three criteria is not fulfilled it is said to be *ill-posed*. All inverse problems in this thesis are ill-posed in one way or another.

To overcome these hurdles, inverse problems are often formulated as a minimization problem

$$u^\sharp \in \operatorname*{argmin}_u \Big\{ \mathcal{D}(Au|f) + \mathcal{P}(u) \Big\}, \tag{1.1}$$

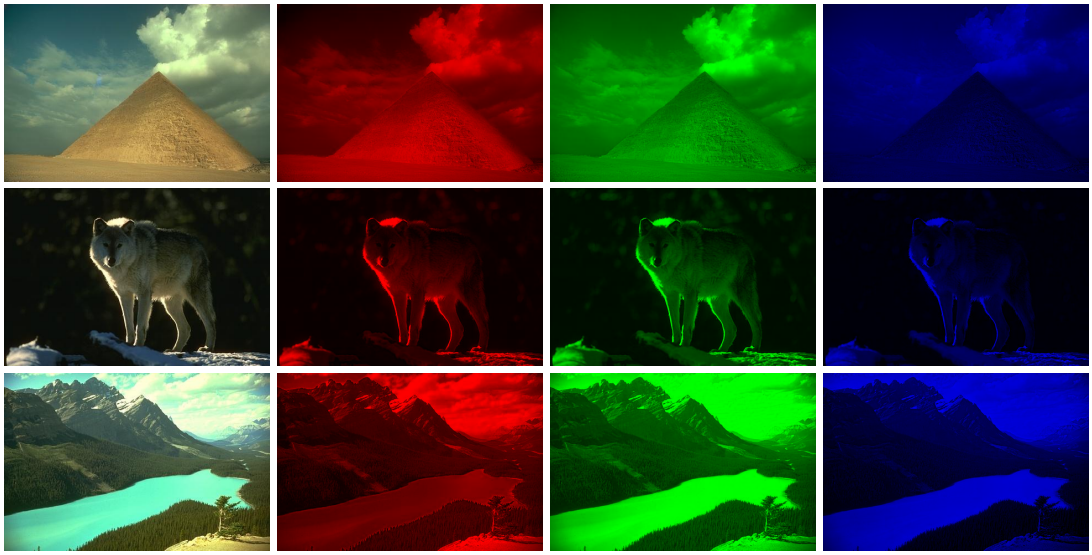


Figure 1.2. *Colour channels of RGB colour images show similar structures.*

where $\mathcal{D}$ is the fidelity between data f and estimated solution Au and $\mathcal{P}$ introduces a priori information about the solution we seek.

A priori information is what we know about the solution we seek before actually seeing it. Therefore, this information is independent of the acquired data. In the context of inverse problems such a priori information can act as regularization – stabilizes the inversion – and is necessary in order to guarantee meaningful solutions. If the modelled system does not uniquely determine a solution – as in compressed sensing [46, 28, 53] – there are multiple candidates that match the data and a priori information can be used to "pick" a favourable candidate.

There is a variety of mathematical models that can be used to formulate a priori knowledge. Commonly, we are seeking a solution of minimal energy [54], a solution which is sparse in some domain [46, 28, 53] or favour a piecewise constant – cartoon-like – solution [120]. However, many multi-modality images show a strong but often unexploited inter-channel correlation. As "function follows structure for the most part" [131] the images to be reconstructed from a PET-MRI or PET-CT scanner are likely to share many edges. Some example images are shown in figure 1.1 where the reconstructions from PET, MRI and CT of a single patient are shown (using standard settings with the scanner software). Despite the fact that all three images show completely different biological information, all three images feature some *similar structures* due to the same underlying anatomy. Similarly in colour photography, all colour channels show the structures of the scenery the data is acquired from, see figure 1.2 for some examples.

This fact has also been observed by [72] on a broader set of images. However, data from different modalities are usually treated independently and reconstructed separately.

1.3. One-Sided Reconstruction.

In medical imaging, the incorporation of structural information from an anatomical image into a functional image has been the subject of research since the early 90's [97] and it is still an active research topic [5, 19, 118, 152, 17, 88, 87, 153, 82]. It has been used in PET-MRI for partial-volume correction, which otherwise is a post-processing step [56]. Similarly in spectral imaging, pan-sharpening is a technique to use a high spatial but low spectral resolution image to enhance images with low spatial but high spectral resolution [7, 111, 110, 59].

Structural information can be incorporated into the reconstruction process

$$u^\sharp \in \operatorname*{argmin}_u \Big\{ \mathcal{D}(Au|f) + \mathcal{P}(u|v) \Big\} \tag{1.2}$$

by the prior $\mathcal{P}$. We call this asymmetric inversion – to estimate $u^\sharp$ using structural information from v – *one-sided reconstruction*. This is also referred to as *model fusion* [140, 118, 19, 17, 88, 74, 86, 7, 111, 110, 59].

A variety of priors have been proposed for one-sided reconstruction and we will discuss some important choices in the following.

Anatomical Edge Weighting and Bowsher Prior. In a discrete set-up the following prior can be formulated

$$\mathcal{P}(u|v) = \sum_{n=1}^{N} \sum_{m \in \mathsf{N}(n)} w_{n,m}(v) f(|u_n - u_m|), \tag{1.3}$$

where the neighbourhood around a pixel n is denoted by $\mathsf{N}(n)$, $w_{n,m}(v)$ is a weight on the relation between pixels n and m and $f\colon [0,\infty) \to \mathbb{R}$. Structural a priori information is represented by the weight $w_{n,m}(v)$ which is set to zero if there is an edge in the anatomical image v between the pixels n and m [97, 41, 33]. As an alternative to this method, which needs an *explicit edge set* from the anatomical image, it has been proposed to weight neighbours according to the similarity in pixel intensity in the anatomical image. This became known as the *Bowsher prior* [19, 88, 17].

Bregman Distance. Another prior that has been proposed for one-sided reconstruction [7, 110, 57, 89, 112] is

$$\mathcal{P}(u|v) = \int_\Omega |\nabla u(x)| - \langle \nabla u(x), \theta_v(x) \rangle \, \mathrm{d}x \tag{1.4}$$

which is closely related to the Bregman distance of total variation [121]. It promotes the alignment of the gradient ∇u to a pre-defined vector field θ_v depending on v. Important to note is that this prior only promotes $\nabla u(x) = s(x)\theta_v(x)$ with $s(x) \geq 0$ and therefore, might not be suitable for all applications. While it has been successfully applied if θ is a normalized gradient from the *same modality* [7, 110, 57, 112] there is no evidence if and why this works in a general multi-modality set up as contrasts can be both positively and negatively correlated between the pair of modalities, see figure 1.1.

Class Modelling. A different branch of incorporating anatomical information is what we call *region based* or *class modelling*. These methods assume that the image consists of a predefined number of regions whose intensities are identically distributed. The task is then to find these regions and the parameters of their distributions. A priori anatomical information can provide information about the shape of the regions [18, 131, 5, 125, 123, 126, 16, 100, 37].

Information Theoretical Measures. A fourth way of exploiting anatomical side information are information theoretical measures such as *joint entropy* and *mutual information* [140, 118, 142, 141] also used in multi-modality image registration [159, 106]. Joint entropy and mutual information favour images whose intensities or gradients are "similarly" distributed – in contrast to all the aforementioned priors which favour local similarities.

1.4. Joint Reconstruction.

One-sided reconstruction can be extended to fully *joint reconstruction* – also called *joint inversion* [74] – as

$$(u^\sharp, v^\sharp) \in \operatorname*{argmin}_{(u,v)} \Big\{ \mathcal{D}(Au|f) + \mathcal{E}(Bv|g) + \mathcal{P}(u,v) \Big\}, \tag{1.5}$$

where we not only seek a solution $u^\sharp$ but a solution pair $(u^\sharp, v^\sharp)$ of the combined problem. Here $\mathcal{D}$ and $\mathcal{E}$ denote the data fidelities and $\mathcal{P}$ the joint prior.

Joint reconstruction of two imaging modalities was to our knowledge first considered by Haber and Oldenburg [76]. While they considered joint reconstruction in a general context, they already mentioned joint reconstruction of function and structure in medical imaging like the combination of PET and MRI. In the early 2000's, joint reconstruction was only considered in geophysics [63, 64, 65, 74]. In 2011, a Bayesian approach to joint inversion for multi-contrast MRI has been proposed by [13]. The different MRI datasets collected in this study show different physiological properties but they all have the anatomy in common. Since 2014, joint reconstruction has also been considered for PET-MRI [92, 94, 52, 51] and multi-energy CT [133]. As we mentioned already in the introduction, joint reconstruction has been used in colour photography since the 90's but became increasingly more popular due to increasing computational resources [135, 22, 80, 50, 112].

Apart from [135, 13, 112], in all of these cases joint reconstruction was cast as a minimization problem of the form (1.5). The priors used in [63, 64, 65, 74, 92, 94, 52, 80, 51, 133] all have in common that they tend to align the image gradients and will be discussed in more detail later in this thesis. Many methods that are suitable for one-sided reconstruction as in (1.2) cannot be used for joint reconstruction with a joint objective function (1.5), e.g. [97, 18, 86, 19, 34, 17, 88, 89]. Information theoretical priors such as joint entropy and mutual information [140, 118, 142, 141] might be suitable for joint reconstruction but none of these priors have been studied yet for joint reconstruction and it is out of the scope of this thesis to carefully consider these as well.

1.5. Contributions.

This thesis is a contribution to *joint reconstruction* in multi-modality imaging. We define a family of priors – named *parallel level sets* – that model the common structure often seen in multi-modality imaging. This general family includes [86, 139, 63] as special cases but reaches beyond that. Its derivation from first principles is one of the novel contributions of this thesis.

We analyse the parallel level set family in terms of convexity and the generated diffusive flow. Although special cases of this family have been considered already, this analysis is the first of its kind for parallel level sets. The convexity analysis not only yields results about this family but we can draw general conclusions about the convexity of priors for multi-modality imaging. The analysis of the diffusive flow gives insights into the parallel level set family from a different perspective. As this analysis holds true

for all members of the family we gain more insights into previously published results [86, 139, 63].

With the help of graphical models, we justify the ad hoc defined functional for joint reconstruction (1.5), cf. [76, 63, 64, 22, 65, 74, 92, 94, 133], and therefore put this approach on solid grounds. In particular, the derivation shows how to weight the fidelity terms from a statistical point of view.

We test the proposed framework in numerous numerical examples in RGB colour imaging and PET-MRI. On the one hand, this is the first attempt to jointly reconstruct PET and MRI. On the other hand, these are the first results to use parallel level sets for medical imaging and RGB demosaicking.

Our results in RGB colour imaging show that the coupling of the colour channels in a joint reconstruction yields superior results in denoising and demosaicking compared to channel-by-channel approaches: we observe sharper images with less colour artefacts.

In addition, we apply joint reconstruction to PET-MRI where the MRI data are severely undersampled. The results show that the PET images are improved despite the undersampling in MRI. Moreover, joint reconstruction of PET-MRI leads to less undersampling artefacts in MRI compared to separate reconstruction.

It is sometimes advantageous in MRI to have a complex image model with magnitude and phase. Under realistic assumptions our numerical experiments on MRI reconstruction show that it is beneficial to formulate the prior knowledge separately on magnitude and phase. This is also necessary in joint reconstruction when complex-valued MRI images are to be combined with real-valued PET images.

The results in this thesis have been published as journal papers [50, 51] and in conference proceedings [52] and have been presented at several national and international conferences.

Peer-Reviewed Journal Papers.

[50] M. J. Ehrhardt and S. R. Arridge. “Vector-Valued Image Processing by Parallel Level Sets”. *IEEE Transactions on Image Processing* 23.1 (2014), pp. 9–18.

[51] M. J. Ehrhardt, K. Thielemans, L. Pizarro, D. Atkinson, S. Ourselin, B. F. Hutton, and S. R. Arridge. “Joint Reconstruction of PET-MRI by exploiting Structural Similarity”. *Inverse Problems* 31 (2015), p. 015001.

Peer-Reviewed Conference Papers.

[52] M. J. Ehrhardt, K. Thielemans, L. Pizarro, P. Markiewicz, D. Atkinson, S. Ourselin, B. F. Hutton, and S. R. Arridge. "Joint Reconstruction of PET-MRI by Parallel Level Sets". *IEEE Nuclear Science Symposium and Medical Imaging Conference*. 2014.

1.6. Outline.

We will begin this thesis with an introduction to PET and MRI in chapter 2 focussed on the essentials for image reconstruction. Following on the deterministic viewpoint in chapter 2, we will accompany the reader in chapter 3 through the statistical modelling of noise for single- and multi-modality imaging as well as image reconstruction from a statistical point of view. Chapter 4 gives an introduction to regularization based on gradient information and shows its connection to diffusion. Chapters 5, 6 and 7 form the theoretical contribution of this thesis. In chapter 5, we will introduce the concept of parallel level sets and propose a variety of priors based on these ideas. These families are analysed with respect to their diffusive flow in chapter 6 and convexity in chapter 7. Numerous numerical examples in colour imaging in chapter 8 and PET-MRI in chapter 9 show the gain of joint reconstruction and in particular of using parallel level sets. We conclude this thesis in chapter 10 and point at possible directions of follow-up research.

Chapter 2

Positron Emission Tomography and Magnetic Resonance Imaging

The two modalities we focus on in medical imaging are positron emission tomography (PET) and magnetic resonance imaging (MRI). In this chapter we will summarize the basics about PET and MRI with a focus on what is needed in the rest of the thesis. Further information, e.g. on physics and clinical applications, can be found in [9, 6] for PET and [99, 108] for MRI. For a detailed mathematical description of both modalities we refer to [55].

2.1. Positron Emission Tomography.

PET is a nuclear medicine and molecular medicine imaging modality where a radioactive labelled tracer is injected into a patient. The tracer is a biologically important molecule that has been radioactively marked with a positron emitting isotope which makes it possible to be monitored. When an emitted positron annihilates with an electron of the tissue, two photons in (almost) opposite directions are sent out and recorded by rings of detectors around the object. The inverse problem in PET then constitutes reconstructing the tracer distribution from a (very large) number of these recorded pairs of photons.

2.1.1. From Photons to Data. Let us denote the expected number of emitted photon pairs as a function of a spatial location by $u\colon \Omega \subset \mathbb{R}^D \to [0, \infty)$, where $D = 2$ or 3. Clearly, u will depend on the tracer concentration, the half-life of the isotope and the acquisition time. We will now explain a simple geometric model for data acquisition. Consider the mth pair of detectors and assume for now that these are infinitely small.

There are two orthogonal vectors $\theta_m, \theta_m^\perp \in \mathbb{R}^D$ such that the line

$$\left\{ s\theta_m + \theta_m^\perp \mid s \in \mathbb{R} \right\} \tag{2.1}$$

passes through both of the detectors. Assuming that the two emitted photons travel perfectly in opposite directions and neglecting physical effects such as scatter or attenuation, then the expected number of photon pairs detected by the mth pair of detectors is given by

$$f_m = c(Xu)_m \coloneqq c \int_{-\infty}^{\infty} u(s\theta_m + \theta_m^\perp)\, \mathrm{d}s, \tag{2.2}$$

with a normalization constant c. The operator X in (2.2) is called the *X-ray transform* [115, 114] which coincides with the *Radon transform* in two dimensions. A practical implementation of this transform would need to take the finite size of the detectors into account – for instance by summing up the integral over several rays that connect the pair of detectors.

2.1.2. Attenuation. So far our PET imaging model only takes geometry into account. We will now extend it to attenuation. Let $u(x)$ be the number of photon pairs emitted at a position x and travelling on the line (2.1). One of the photons is travelling on a half line or ray $\{s\theta_m + x \mid s \geq 0\}$ with origin x and direction θ_m and the other one in direction $-\theta_m$. Let us consider for the moment only one of these photons. *Beer's Law* says that the reduction of photons due to attenuation on a (small) distance Δs is proportional to Δs and $u(x)$, i.e.

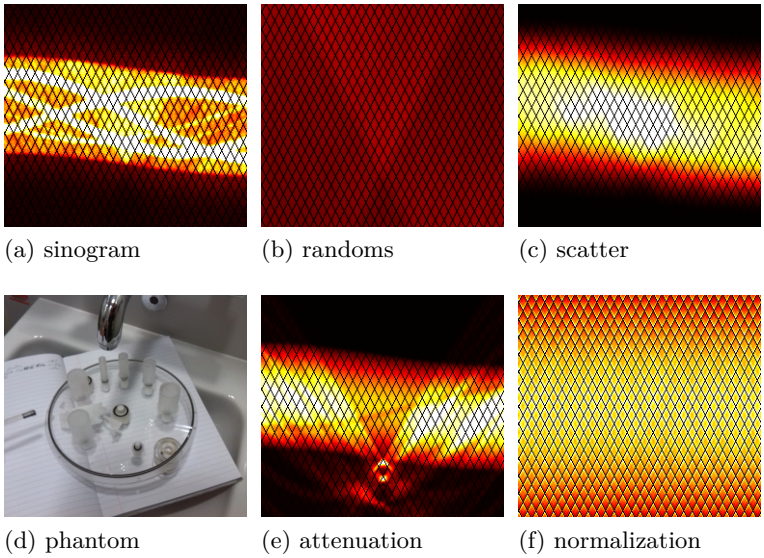
$$u(\Delta s\theta_m + x) - u(x) = -\mu(\Delta s\theta_m + x)u(x)\ \Delta s, \tag{2.3}$$

where we call the proportionality factor $\mu(x)$ the *attenuation coefficient* at x. The limiting ordinary differential equation for $\Delta s \to 0$

$$\partial_{\theta_m} u(x) = -\mu(x)u(x) \tag{2.4}$$

is solved by

$$u(s\theta_m + x) = u(x) \exp\left(- \int_0^s \mu(t\theta_m + x)\, \mathrm{d}t \right). \tag{2.5}$$



(a) sinogram (b) randoms (c) scatter

(d) phantom (e) attenuation (f) normalization

Figure 2.1. *Data in PET (a) of a phantom scan (d) with correction factors for several physical effects that are not accounted for in simple geometric modelling (b,c,e,f). The range of the colour maps are (a) [0,450], (b,c) [0,100], (e) [1,12] and (f) [0.075,0.18].*

Therefore, in terms of pairs of detected photons travelling in opposite directions – the pair gets attenuated over the whole line – outside the object we expect to observe

$$\begin{aligned} &u(x)\exp\left(-\int_0^\infty \mu(t\theta_m + x)\,\mathrm{d}t\right)\exp\left(-\int_{-\infty}^0 \mu(t\theta_m + x)\,\mathrm{d}t\right) \\ &\quad = u(x)\exp\left(-\int_{-\infty}^\infty \mu(t\theta_m + x)\,\mathrm{d}t\right). \end{aligned} \tag{2.6}$$

Note that the loss of photons emitted at $x = s\theta_m + \theta_m^\perp$ is independent of s as we integrate over the whole line, that is

$$\int_{-\infty}^\infty \mu(t\theta_m + x)\,\mathrm{d}t = \int_{-\infty}^\infty \mu((s+t)\theta_m + \theta_m^\perp)\,\mathrm{d}t \tag{2.7}$$

$$= \int_{-\infty}^\infty \mu(t\theta_m + \theta_m^\perp)\,\mathrm{d}t = (X\mu)_m. \tag{2.8}$$

We will denote the *attenuation factors* by $\mu_m = \exp[(X\mu)_m]$. Finally, including attenuation into the previous geometric model leads to

$$f_m = \frac{c}{\mu_m}(Xu)_m. \tag{2.9}$$

2.1.3. PET Forward Model and Correction Factors. In the previous section we have seen that we can compute a data estimate by means of a weighted X-ray transform depending linearly on the tracer concentration. This simplified model assumes that all

photons are travelling on straight lines and that a pair of photons is always detected. As such the model ignores important physical effects like scatter and randoms. When a photon gets scattered it changes its direction. It might still be detected but the annihilation did not take place along the line between the two detected photons. Similarly, if two photons are detected it can happen that they are being associated as a pair although they were not emitted by the same event. This may arise if two events occur close in time and when one photon of each pair does not get detected – for instance it might have been attenuated or gone out of the scanner without a detector in its way.

These effects are usually incorporated by additive correction factors that are estimated based on the measured data. In addition, *normalization factors* are incorporated to correct for detector efficiency such that the data of uniform phantoms get reconstructed uniformly. Calling the vectors of scatter correction factors ξ, random correction factors ζ, normalizations factors ν and attenuation factors μ, the PET forward operator becomes

$$A \cdot + b \colon \mathbb{R}^N \to \mathbb{R}^M, \quad (Au + b)_m \coloneqq \frac{c}{\nu_m \mu_m} (Xu)_m + b_m \tag{2.10}$$

with background $b \coloneqq \zeta + \xi$. All these factors for a real phantom study are visualized in figure 2.1. As is it can be seen, some of these factors are object and some are scanner / geometry dependent. For instance, the triangular pattern – one big triangle – that can be seen in the estimation of the randoms and attenuation originates from the bed in the scanner. The chequered pattern in all images is due to gaps in the detector ring and when properly modelled does not influence the reconstruction at all.

2.2. Magnetic Resonance Imaging.

Magnetic Resonance Imaging (MRI) is a non-ionizing, non-invasive, in-vivo imaging technique with a wide range of applications. In this section we will give a very brief description of the underlying physics and state the imaging equation. Based on this we will describe some more advanced topics that are needed in this thesis.

2.2.1. From Protons to Data. MRI exploits the spin of hydrogen protons in the human body. In MRI the patient is exposed to a very strong magnetic field whose strength is – for a modern 3 Tesla scanner – around 60,000 times larger than the earth's magnetic field [108]. This magnetic field makes the protons' spins rotate with the so called Larmor frequency ω_0 around the scanner's axis and creates a net magnetic vector.

When a radio frequency (RF) pulse is applied, the magnetic vector gets deflected. Switching off the RF pulse allows the magnetic vector to realign with the main magnetic field. The relaxation of the magnetic vector and the axial spin follow an exponentially decay with time constants T_1 and T_2. Therefore, depending on the time when we acquire the signal we might get a signal that includes more of the effects of T_1 and sometimes more of T_2. The contrast in MRI comes from differences in T_1 and T_2 of different types of tissue.

Independent of the imaging set-up, we will denote the MRI image by $v\colon \Omega \to \mathbb{R}$. Then the imaging equation MRI comes down to is

$$g(t) \approx c_1 e^{\mathtt{i}\omega_0 t} \int_\Omega v(x)\, s(x)\, e^{-t/T_2(x)}\, e^{-2\pi \mathtt{i}\langle x, k(t)\rangle}\, \mathrm{d}x \tag{2.11}$$

$$= c_1 e^{\mathtt{i}\omega_0 t} \int_\Omega \rho(x) e^{-2\pi \mathtt{i}\langle x, k(t)\rangle}\, \mathrm{d}x \tag{2.12}$$

which is a scaled *Fourier transform* of ρ – the product of the MRI image v, the spatially varying receiver sensitivity s and the T_2 decay e^{-t/T_2}. The brackets $\langle x, y\rangle \coloneqq \sum_n x_n y_n$ denote the Euclidean inner product. The application of one RF pulse allows us to acquire a sequence of data $g(t_0), g(t_1), \ldots, g(t_M)$ until the signal has decayed due to $\exp(-t/T_2)$.

The collected data are Fourier samples at spatial frequencies $k(t_0), k(t_1), \ldots, k(t_M)$ with

$$k\colon [0, T) \to \mathbb{R}^D, \qquad k(t) = c_2 \int_0^t G(\tau)\, \mathrm{d}\tau \tag{2.13}$$

where the time is normalized so that the RF pulse is applied at time zero and G is the gradient generated from the gradient coils. It is this symbol that gives the data domain in MRI the name *k-space*. We have control over the sampling as we can switch on and off the gradient $G(\tau) \in \mathbb{R}^D$. There are physical constraints on the maximal absolute values of each component $|G_d|$ but these are not very important for our considerations. More important is that k traverses the k-space along a continuous curve for each RF pulse and sampling along this curve is not time limiting.

2.2.2. Acquisition Time. Due to the sequential acquisition of the data, the scan time can be quite long. Speeding up the acquisition is a main research topic in MRI for several reasons: A faster acquisition is beneficial to save money by increasing the patient through-put and to enhance patient's comfort as the stay in the scanner is shortened.

Maybe more importantly, for some applications such as monitoring dynamic processes [66] or paediatric imaging [151] a short data acquisition is crucial to obtain meaningful data.

A peculiarity of MRI is that (spatial) resolution and scanning time are connected. The resolution in MRI can be enhanced by taking more and more measurements. Again, as the acquisition is inherently sequential, there is a trade-off between spatial resolution and acquisition time.

Proposed solutions to speed up the acquisition are to acquire more data at the same time, parallel MRI, or to omit redundant measurements, sparse MRI. We will discuss these two techniques later in this section after a formal introduction of the components of the MRI forward model.

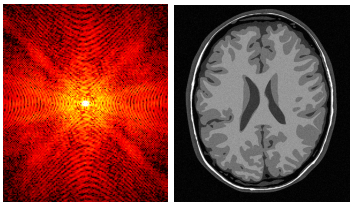
2.2.3. Sampling. As we have seen, acquiring MRI data can be done in many different ways using the same scanner. There is a large variety of sampling schemes like Cartesian (line-by-line), radial or spiral sampling [104]. They all have in common that they are acquired on a finite number of continuous trajectories in k-space.

In formulas let $\mathsf{Z} \in \{1, \dots, N\}^M$ be a vector of indices between 1 and N counting all the data measured by the MRI scanner. We define a sampling operator Z^* which mimics this acquisition as the linear operator $Z^* \colon \mathbb{C}^N \to \mathbb{C}^M, (Z^* f)_m = f_{\mathsf{Z}[m]}$. To avoid double subscripts and so to enhance readability we will denote the nth component of a vector sometimes by x_n and sometimes by $x[n]$. The case of interest is, of course, when we acquire a lot less samples than we would usually need to uniquely identify our image, i.e. $M \ll N$.

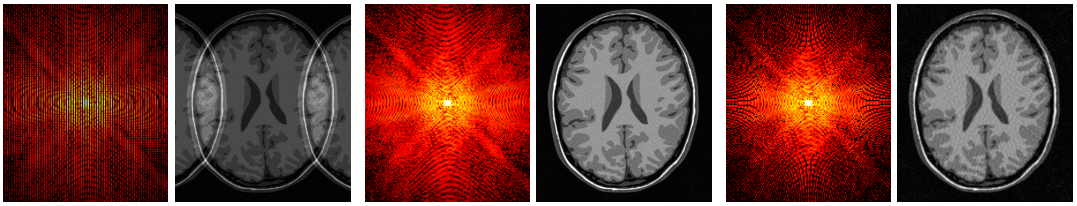
Some standard sampling schemes like equidistant Cartesian, radial and spiral sampling are depicted in figure 2.2 (a,b), (c-e) and (f-g), respectively. When we assume that the image of interest is only real valued, acquiring half the k-space can be sufficient, cf. figure 2.2 (i,j). Real images have a symmetric k-space – point symmetry to the origin with complex conjugate – so that we can reconstruct the image perfectly from one "half" of the k-space when this symmetry is exploited.

For reconstruction we need the adjoint of the sampling operator Z^*.

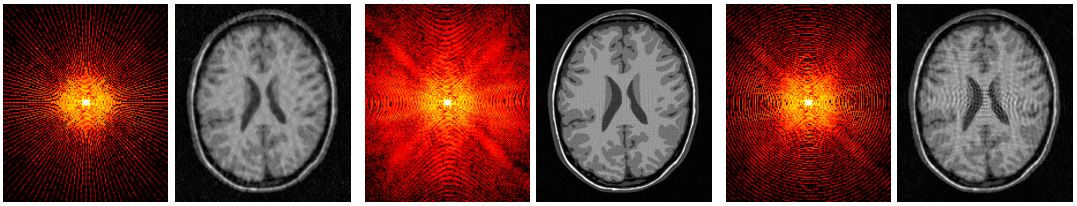
Definition 2.2.1. *Let* X, Y *be inner product spaces and denote their inner products by* $\langle \cdot, \cdot \rangle_{\mathsf{X}}, \langle \cdot, \cdot \rangle_{\mathsf{Y}}$. *Furthermore, let* $A \colon \mathsf{X} \to \mathsf{Y}$ *be linear. The* adjoint *of* A *– denoted* A^*



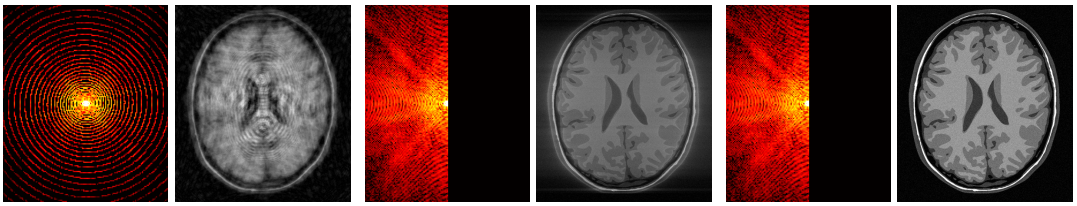
(a) full sampling, 100%



(b) every other line, 50% (c) radial, 40% (d) radial, 22%



(e) radial, 11% (f) spiral, 43% (g) spiral, 23%



(h) spiral, 11% (i) half Fourier, 50% (j) half Fourier – corrected, 50%

Figure 2.2. *Different data samplings (only the central part is shown) and minimal norm reconstructions in MRI. The number indicates the percentage of sampling compared to full sampling. Half Fourier sampling (i,j) can be used to speed up the acquisition when the image is real-valued but the missing data needs to be calculated first to get a good reconstruction.*

– *is defined implicitly such that for all* $x \in \mathsf{X}$ *and* $y \in \mathsf{Y}$ *there is*

$$\langle Ax, y\rangle_{\mathsf{Y}} = \langle x, A^* y\rangle_{\mathsf{X}}\,. \tag{2.14}$$

For ease of notation we will use the *Kronecker delta* $\delta_{n,m}$ defined as

$$\delta_{n,m} \coloneqq \begin{cases} 1, & \text{if } n = m \\ 0, & \text{else} \end{cases}\,. \tag{2.15}$$

Lemma 2.2.2. *The adjoint of the sampling operator Z^* is given by*

$$(Z^*)^* = Z \colon \mathbb{C}^M \to \mathbb{C}^N, \qquad (Zg)_n \coloneqq \sum_{m=1}^{M} g_m \delta_{n,\mathsf{Z}[m]}. \tag{2.16}$$

Proof. We use the standard inner product on $\mathbb{C}^K$, i.e. $\langle x, y\rangle_{\mathbb{C}^K} \coloneqq \sum_{k=1}^{K} x_k \overline{y_k}$ with $\overline{y}$ denoting the complex conjugate of y.

Let $f \in \mathbb{C}^N, g \in \mathbb{C}^M$ and $\mathsf{Z} \in \{1, \dots, N\}^M$. Then there is

$$\langle f, Zg\rangle_{\mathbb{C}^N} = \sum_{n=1}^{N} f_n \overline{(Zg)_n} = \sum_{n=1}^{N} f_n \sum_{m=1}^{M} \overline{g_m} \delta_{n,\mathsf{Z}[m]} = \sum_{m=1}^{M} \left(\sum_{n=1}^{N} f_n \delta_{n,\mathsf{Z}[m]} \right) \overline{g_m} \tag{2.17}$$

$$= \sum_{m=1}^{M} f_{\mathsf{Z}[m]} \overline{g_m} = \sum_{m=1}^{M} (Z^* f)_m \overline{g_m} = \langle Z^* f, g\rangle_{\mathbb{C}^M}. \tag{2.18}$$

■

The sampling and its adjoint were defined in very general and abstract terms. We shall make these definitions clear with the following example.

Example 2.2.3. Let $f \in \mathbb{C}^3, g \in \mathbb{C}^4$ and $\mathsf{Z} = (2, 1, 2, 2) \in \{1, 2, 3\}^4$. The sampling of f at Z is given by

$$Z^* f = (f_{\mathsf{Z}[1]}, f_{\mathsf{Z}[2]}, f_{\mathsf{Z}[3]}, f_{\mathsf{Z}[4]}) = (f_2, f_1, f_2, f_2) \tag{2.19}$$

and its adjoint by

$$Zg = \left(\sum_{m=1}^{4} g_m \delta_{1,\mathsf{Z}[m]}, \sum_{m=1}^{4} g_m \delta_{2,\mathsf{Z}[m]}, \sum_{m=1}^{4} g_m \delta_{3,\mathsf{Z}[m]} \right) = (g_2, g_1 + g_3 + g_4, 0). \tag{2.20}$$

▲

Remark 2.2.4. If every data point is sampled at most once, i.e. the vector Z contains every index at most once, then the operator Z performs a zero-filling operation – the missing data is set to zero – which explains our notation. ▲

2.2.4. MRI operator. In this part we will define the forward model that is used to reconstruct data in MRI. The MRI image v is connected to the data g by the MRI operator $B \colon \mathbb{R}^N \to \mathbb{R}^{2M}$ which is related to the Fourier transform. To be more precise the operator B is given by

$$B \colon \mathbb{R}^N \overset{\mathrm{Re}^*}{\to} \mathbb{R}^{2N} \simeq \mathbb{C}^N \overset{F}{\to} \mathbb{C}^N \overset{Z^*}{\to} \mathbb{C}^M \simeq \mathbb{R}^{2M}, \tag{2.21}$$

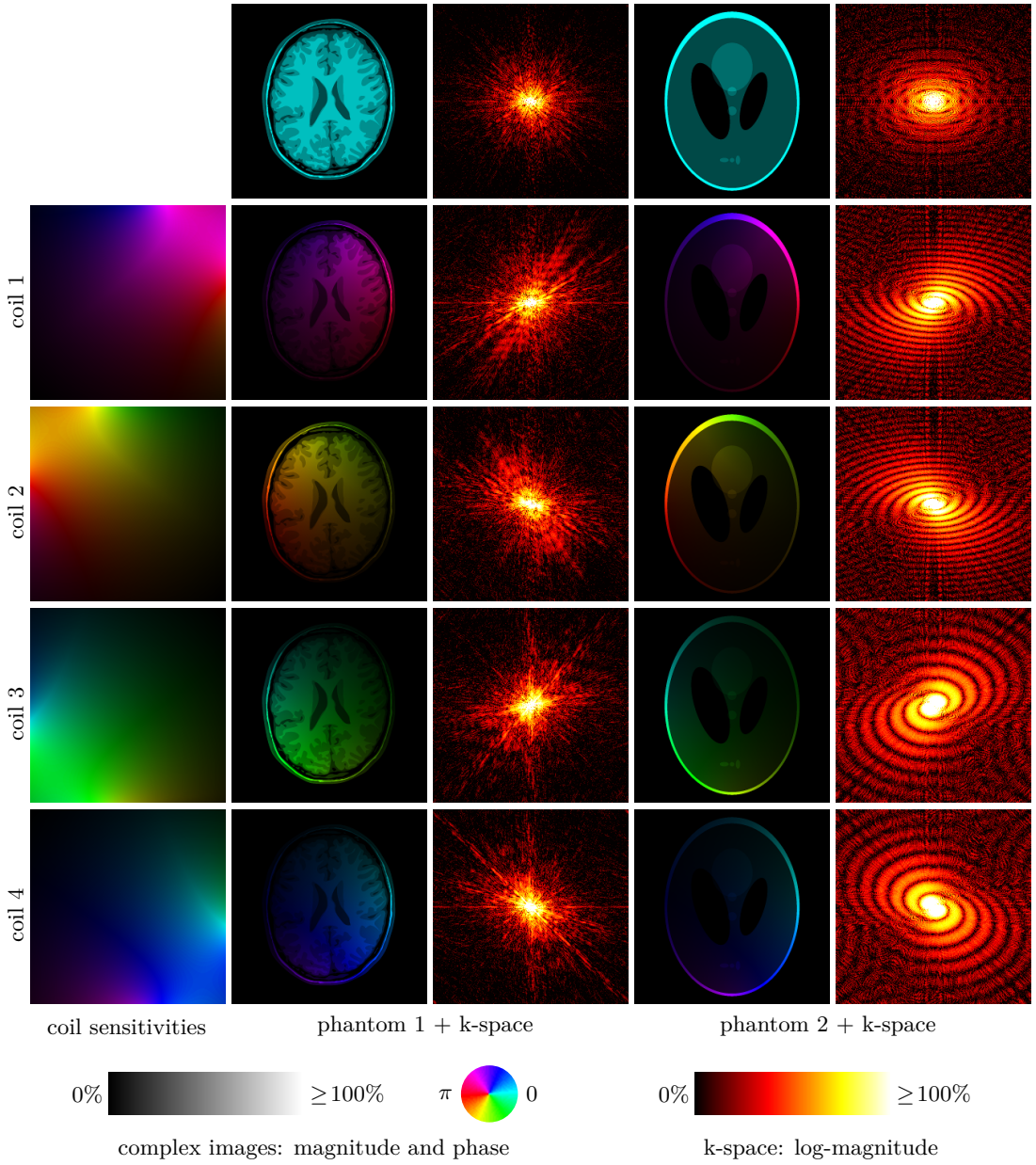


Figure 2.3. *In parallel MRI – here with four receiver coils – the acquired data are samples from the Fourier transform (k-space) spatially weighted by complex coil sensitivities shown on the left. On the right are the weighted phantoms and the magnitude of their Fourier transforms. All the complex images are shown with magnitude encoded in brightness and phase in colours.*

i.e. we identify our real-valued image with a complex image and then sample from its Fourier transform F.

We denoted the operator that embeds a real-valued image into complex space by $\mathrm{Re}^*\colon \mathbb{R}^N \to \mathbb{R}^{2N} \simeq \mathbb{C}^N$ with $\mathrm{Re}^*(x) = (x, 0) \simeq x$. It is the adjoint operator of the "reverse" operator that takes the real part of a complex number, i.e. $\mathrm{Re}\colon \mathbb{C}^N \simeq \mathbb{R}^{2N} \to \mathbb{R}^N$ with $\mathrm{Re}(x + \mathtt{i}y) = x$ for $x, y \in \mathbb{R}^N$. It is important to note that these operators are indeed adjoint if the inner product on $\mathbb{C}^N$ is chosen to be $\mathrm{Re}\,\langle x, y\rangle_{\mathbb{C}^N}$.

Therefore, there is a simple reconstruction scheme for all kind of MRI data g. First, for each measured k-space location either average all the samples at this point or pick a single sample. Then there is a unique image

$$v^\sharp = B^* g = \operatorname{Re} F^{-1} Z g \tag{2.22}$$

that has the smallest norm among all images that match the data perfectly. It is called the *minimal norm* solution or in the context of MRI *zero-filled* solution. It is obtained by setting all missing data to zero and taking the real part of the inverse Fourier transform. As we can see in figure 2.2 for moderate amount of undersampling this gives reasonable results but for more undersampling, artefacts can be observed that make some of these images unusable.

2.2.5. Parallel MRI. An option to reduce acquisition time in MRI is to collect more data simultaneously with several receiver coils. These receiver coils have spatially varying sensitivities and are therefore measuring different parts of the object [130, 79, 96]. We will formulate the reconstruction of parallel MRI as a linear inverse problem using explicit sensitivity maps which became known as SENSE (SENSitivity Encoding) [130]. There are other methods working in the k-space like GRAPPA [69] or a combination of these two [148].

With an abuse of notation we denote the parallel MRI forward operator again by $B\colon \mathbb{R}^N \to \mathbb{R}^{2MK}$ which maps an image $v \in \mathbb{R}^N$ to the data collected by the K receiver coils $g \in \mathbb{R}^{2M} \times \ldots \times \mathbb{R}^{2M}$ with $(Bv)_k = Z^* F S_k \operatorname{Re}^* v$ with (assumed to be) known coil sensitivities $s_k \in \mathbb{C}^N$ and $S_k \coloneqq \operatorname{diag}(s_k)$.

$$B\colon \mathbb{R}^N \xrightarrow{\operatorname{Re}^*} \mathbb{C}^N \xrightarrow{S_k} \mathbb{C}^{NK} \xrightarrow{F} \mathbb{C}^{NK} \xrightarrow{Z^*} \mathbb{C}^{MK} \tag{2.23}$$

Two simulated parallel MRI experiments with four receiver coils are shown in figure 2.3. From left to right are the complex sensitivity maps, the weighted images and their Fourier transforms without undersampling. It is important to see that the sampling in parallel MRI is the same for *all* receiver coils. This definition is a generalization of (2.21) as they coincide for a single coil with homogeneous sensitivity.

2.2.6. Sparse MRI. Reconstruction with less data by incorporating prior information in the form of sparsity became known as *compressed* or *compressive sensing* [27,

29, 30, 46, 53] and was applied to MRI reconstruction, e.g. [35, 101, 104, 149, 93, 132, 13]. It has been shown that a priori information can be used to decrease the amount of data needed to reconstruct an image of similar quality. Successfully exploited a priori knowledge in MRI includes, for instance, sparsity in the gradient [134] or wavelet domain [104], sparsity in a self-learned dictionary [132] or similarity from different MRI contrasts [13]. In this thesis, the additional a priori knowledge is the similarity of the MRI image to the simultaneously acquired PET image.

2.3. Summary.

This chapter served as an introduction to PET and MRI. We derived the forward models that are a necessary ingredient in the reconstruction of the images. Data acquisition time in MRI is a major limitation and techniques that can handle undersampled k-space like parallel MRI and sparse MRI have been proposed. While in parallel MRI extra information comes from several receiver coils, sparse MRI incorporates a priori information. In a PET-MRI scanner, extra a priori knowledge might come from the simultaneously acquired PET.

Chapter 3

Statistical Modelling for Multi-Modality Imaging

In this chapter we will have a closer look at the objective function for joint minimization. In the introduction we stated one way of combining two inverse problems but one can argue that this formulation is somewhat ad hoc. In the following we will have a detailed look at the statistical modelling that forms the basis of image reconstruction and extend this to multi-modality imaging. Graphical models enable a statistically sound derivation of a joint objective function like (1.5) which has been used ad hoc [76, 63, 64, 22, 65, 74, 92, 94, 133].

3.1. Data Acquisition and Noise Modelling.

Our description of the one modality case is adapted from [12, 23]. We consider an imaging system with forward operator A which maps an image u to the data f as we have seen in the previous chapter for PET and MRI. The *first assumption* is that the data is discretized, i.e. given as a vector $f = (f_m)_{m=1}^M \in \mathbb{R}^M$ and is a realization of the vector-valued random variable $\mathtt{F} = (\mathtt{F}_m)_{m=1}^M$. The object of interest u is described by either a function or a vector – there is no need to specify this at the moment. Therefore, the imaging model $A\colon \mathsf{X} \to \mathbb{R}^M$ is either a semi-discrete or discrete operator, with X denoting a generic function space. The probability (density) of interest is the conditional probability $p(f|u)$ – the probability of observing f given u. The further modelling is based on *two more assumptions*. First, we assume that any pair of random variables $\mathtt{F}_n$, $\mathtt{F}_m$ associated to different detector elements are conditionally independent given the image u, i.e. $p(f|u) = \prod_{m=1}^M p(f_m|u)$. Moreover, the expected value of $\mathtt{F}_m$ is given by $(Au+b)_m$, i.e. $\mathbb{E}(\mathtt{F}_m) = (Au+b)_m$, with constant offset b_m taking into account effects such as scatter or randoms in the case of PET, see the previous chapter. This relates the image of interest to the data. To complete this paragraph, we summarize

the assumptions we made.

1) discrete data: $$f = (f_m)_{m=1}^M \in \mathbb{R}^M \tag{3.1}$$

2) conditional independence: $$p(f|u) = \prod_{m=1}^{M} p(f_m|u) \tag{3.2}$$

3) expectation: $$\mathbb{E}(\mathtt{F}_m) = (Au + b)_m \qquad m = 1, \dots, M \tag{3.3}$$

Remark 3.1.1. The assumptions (3.1)-(3.3) are sufficient but could be relaxed. For instance one could include continuous data [98, 90] and correlated noise [85] which would make the description more involved. ▲

The aforementioned set-up is very general and holds for a wide range of modalities and different noise distributions. In this thesis we will consider only Gaussian and Poisson noise.

Example 3.1.2 (Noise Models for Imaging). The most common noise model in imaging is *Gaussian* noise. Also the noise in MRI is commonly modelled as additive Gaussian with standard deviation depending on, among others, the temperature, the receiver bandwidth and the field strength [70, 105, 44]. That means that we acquire data $\mathtt{F}_m \sim \text{Normal}(\mu_m, \sigma_m^2)$ with probability density function

$$p(f_m|\mu_m, \sigma_m^2) = \frac{1}{\sigma_m\sqrt{2\pi}} \exp\left(-\frac{(f_m - \mu_m)^2}{2\sigma_m^2}\right), \tag{3.4}$$

expected value $\mu_m \in \mathbb{R}$ and variance $\sigma_m^2 > 0$. Roughly speaking, the probability of observing an instance f_m of the data random variable $\mathtt{F}_m$ is $p(f_m|\mu_m, \sigma_m^2)$. A precise explanation would need to involve intervals, as points of a continuous random variable have always probability zero.

The data in PET is the number of detected photon pairs and as such intrinsically integer-valued. The noise in PET is commonly modelled to be *Poisson* [119, 6, 12]. In formulae, $\mathtt{F}_m \sim \text{Poisson}(\lambda_m)$ with the probability of measuring f_m being

$$p(f_m|\lambda_m) = \exp(-\lambda_m)\frac{\lambda_m^{f_m}}{f_m!}, \tag{3.5}$$

where both the expected value and the variance are $\lambda_m \geq 0$. ▲

Remark 3.1.3. There is sometimes confusion about the noise model in MRI. If we do not consider MRI reconstruction iteratively but just inverting the Fourier transform,

then the noise in the magnitude images follows a Rician distribution. This results from an assumed Gaussian noise of the real and imaginary component in the signal space [70, 105]. But as we do consider reconstruction of images and not the post-processing after Fourier inversion, the Rician distribution is of no importance in this thesis. ▲

3.2. Maximum Likelihood Estimate.

With these assumptions on an imaging system, we can estimate images $u^\sharp$ for some acquired data f. One option is to estimate an image $u^\sharp$ such that the observation of the data f was most likely – the *maximum likelihood* estimate. In formulae, we seek an estimate $u^\sharp$ that maximizes the conditional probability $p(f|u)$

$$u^\sharp \in \operatorname*{argmax}_u p(f|u) \tag{3.6}$$

– if such a maximum exists. This is equivalent to minimizing the negative logarithm – negative log-likelihood –

$$u^\sharp \in \operatorname*{argmin}_u \{-\log p(f|u)\}. \tag{3.7}$$

Example 3.2.1 (Likelihood for Gaussian Noise, MRI). Without background – the standard model for MRI – and Gaussian noise, i.e. $\mathtt{G}_m \sim \text{Normal}((Bv)_m, \sigma_m^2)$, the likelihood is given by

$$p(g|v) \overset{(3.2)}{=} \prod_{m=1}^{M} p(g_m|(Bv)_m, \sigma_m^2) \overset{(3.4)}{=} \prod_{m=1}^{M} \left[\frac{1}{\sigma_m\sqrt{2\pi}} \exp\left(-\frac{(Bv-g)_m^2}{2\sigma_m^2}\right)\right] \tag{3.8}$$

which – assuming white noise, i.e. $\sigma_m \equiv \sigma$ – results in

$$-\log p(g|v) = -\log\left\{\prod_{m=1}^{M}\left[\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(Bv-g)_m^2}{2\sigma^2}\right)\right]\right\} \tag{3.9}$$

$$= \sum_{m=1}^{M}\left[\log(\sigma\sqrt{2\pi}) + \frac{1}{2\sigma^2}(Bv-g)_m^2\right] \tag{3.10}$$

$$= M\log(\sigma\sqrt{2\pi}) + \frac{1}{2\sigma^2}\left|Bv-g\right|^2, \tag{3.11}$$

with the Euclidean norm $|x|^2 \coloneqq \sum_n |x_n|^2$. Equivalently, we could also minimize – what we will call the *MRI data fidelity* –

$$\mathcal{E}(Bv|g) \coloneqq \frac{1}{2\sigma^2}\left|Bv-g\right|_2^2. \tag{3.12}$$

▲

Example 3.2.2 (Likelihood for Poisson Noise, PET). In the case of Poisson noise, i.e. $\mathtt{F}_m \sim \text{Poisson}((Au+b)_m)$, the negative log-likelihood is given by

$$-\log p(f|u) \overset{(3.2)}{=} -\log\left[\prod_{m=1}^{M} p(f_m|(Au+b)_m)\right] \tag{3.13}$$

$$\overset{(3.5)}{=} \sum_{m=1}^{M}\left\{-\log\left[\exp(-(Au+b)_m)\frac{(Au+b)_m^{f_m}}{f_m!}\right]\right\} \tag{3.14}$$

$$= \sum_{m=1}^{M}\left\{(Au+b)_m - f_m\log(Au+b)_m + \log(f_m!)\right\}. \tag{3.15}$$

Adding terms independent of u and therefore equivalent in the minimization, we can equivalently minimize the *Kullback-Leibler divergence* of $Au+b$ and f

$$\mathcal{D}(Au+b|f) \coloneqq \sum_{m=1}^{M}\left\{f_m\log\left[\frac{f_m}{(Au+b)_m}\right] + (Au+b-f)_m\right\}. \tag{3.16}$$

We will refer to this also as *PET data fidelity*. The Kullback-Leibler divergence is non-negative and zero if and only if $Au+b=f$, and therefore suitable to measure the distance between the estimated data $Au+b$ and the measured data f.

This form differs from the statistical version of the Kullback-Leibler divergence for distributions by the term $(Au+b-f)_m$. When $Au+b$ and f sum to one these two versions are the same. ▲

3.3. Extension to Multi-Modality Imaging.

So far we have treated the modalities separately. We will discuss now how they can be linked. We formulate our beliefs using probabilistic modelling and graphical models [14, 8] assuming that both u and v are random variables. Our intuitive belief about these unknown images u, v are described in the graphical model shown in figure 3.1. On the one hand, both u and v depend on a common object and are therefore not independent of each other. The data are modelled as random variables with expectation depending on these images, hence, also the data of f and g are not independent. On the other hand, if we have knowledge about u, then v does not provide extra information about f and vice versa. This means, that f and g are conditionally independent given u and v. Formally, this leads to a separation of the multi-modality likelihood

$$p(f,g|u,v) = p(f|u,v)p(g|u,v) = p(f|u)p(g|v), \tag{3.17}$$

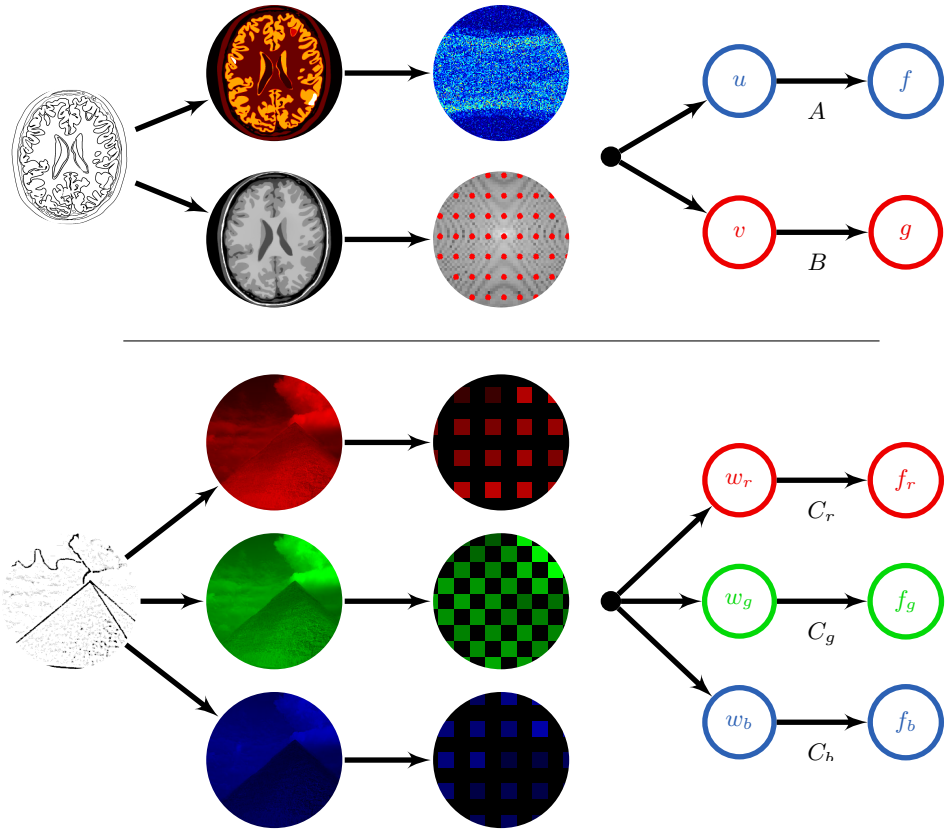


Figure 3.1. *Graphical models for multi-modality imaging. In the top row the graphical models (left in terms of images and right in formulas) show our model for joint PET-MRI reconstruction with simultaneous acquisition. The PET and MRI images u, v are "linked" indirectly via the common anatomy. Similarly, in the bottom row, it can be extended to the case of three modalities as in RGB colour imaging.*

where we used that the data are conditionally independent given the images in the first equation. The second equation follows from the conditional independence of u and g given v and the other way around.

Hence, we can set up a joint maximum likelihood approach by

$$(u^\sharp, v^\sharp) \in \operatorname*{argmin}_{(u,v)} \{-\log p(f|u) - \log p(g|v)\} \tag{3.18}$$

or similarly for three modalities

$$(w_r^\sharp, w_g^\sharp, w_b^\sharp) \in \operatorname*{argmin}_{(w_r,w_g,w_b)} \{-\log p(f_r|w_r) - \log p(f_g|w_g) - \log p(f_b|w_b)\}. \tag{3.19}$$

It is important to see that in this formulation the estimates $(u^\sharp, v^\sharp)$ – or $(w_r^\sharp, w_g^\sharp, w_b^\sharp)$ for RGB imaging – are still independent of one another.

3.4. Maximum A Posteriori Estimator.

Instead of estimating the images so that the observed data is most likely, we can estimate images that are themselves most likely for the acquired data. This leads to *maximum a posteriori* estimators which maximize the *posterior probability* of an image given the observed data $p(u|f)$. Maximizing the posterior probability is equivalent to minimizing the negative logarithm of the posterior probability, i.e. we are seeking a minimum of

$$-\log p(u|f) = -\log\left[\frac{p(f|u)p(u)}{p(f)}\right] = -\log p(f|u) - \log p(u) + \log p(f), \tag{3.20}$$

where the first equality is due to Bayes' formula. The terms $p(u)$ and $p(f)$ are the (marginal) probabilities for an image u and data f. In particular, $p(u)$ is the *a priori probability* of an image u independent of any particular measured data.

Example 3.4.1 (Gibbs Prior). Most prior knowledge can be well formulated as a cost function $\mathcal{P}$ such that favourable images have lower cost. This knowledge can be incorporated into a maximum a posteriori estimate by a Gibbs prior $p(u) = c\exp(-\alpha\mathcal{P}(u))$ with a normalization constant c. Then, maximum a posteriori estimation is equivalent to the minimization problem

$$u^\sharp \in \operatorname*{argmin}_u \{\mathcal{J}(u) = -\log p(f|u) + \alpha\mathcal{P}(u)\}. \tag{3.21}$$

Note, that this coincides with the minimization formulation to solve an inverse problem, cf. (1.1). ▲

Remark 3.4.2. The prior states the a priori probability of u before having seen any data. When $\int \exp(-\alpha\mathcal{P}(u))\,\mathrm{d}u = \infty$ then $p(u)$ this is not a well-defined probability distribution and the formal derivation in the probabilistic sense is not valid any more. In these cases one can use the classical inverse problems theory [54] and think of $\mathcal{P}$ as a cost functional which has high cost / large values if u is in an undesirable state. ▲

Example 3.4.3 (Maximum A Posteriori Estimate for a Multi-Modality System). In the framework of PET-MRI a maximum a posteriori estimate is given by a minimum of

$$\mathcal{J}(u,v) = -\log p(f|u) - \log p(g|v) + \alpha\mathcal{P}(u,v) \tag{3.22}$$

$$= \mathcal{D}(Au+b|f) + \mathcal{E}(Bv|g) + \alpha\mathcal{P}(u,v), \tag{3.23}$$

where we used the separation of the likelihood (3.17). It is very important to note that in this statistical setting the factor $(2\sigma^2)^{-1}$ in $\mathcal{E}$ is a natural weighting of the data fidelity terms, cf. (3.12).

Similarly in RGB imaging the joint objective function becomes

$$\mathcal{J}(w_r, w_g, w_b) = \sum_{n\in\{r,g,b\}} \Big\{ -\log p(f_n|w_n) \Big\} + \alpha\mathcal{P}(w_r, w_g, w_b) \tag{3.24}$$

$$= \sum_{n\in\{r,g,b\}} \Big\{ \frac{1}{2\sigma_n^2} |C_n w_n - f_n|^2 \Big\} + \alpha\mathcal{P}(w_r, w_g, w_b), \tag{3.25}$$

where we assumed uncorrelated, white Gaussian noise in all three colour channels.

Note that this is a statistically rigorous derivation of the joint functional (1.5) which has been used ad hoc for joint reconstruction [76, 63, 64, 22, 65, 74, 92, 94, 133]. ▲

3.5. Derivatives of the Data Fidelities.

In this section we compute the derivatives of the data fidelities for Gaussian and Poisson noise and, in particular, have a look at the gradient of PET, MRI and parallel MRI. These results are very important as we want to perform gradient (derivative) based optimization.

Lemma 3.5.1. *Let $A \in \mathsf{L}(\mathbb{R}^N, \mathbb{R}^M)$, i.e. a linear mapping between $\mathbb{R}^N$ and $\mathbb{R}^M$, and $\mathcal{E}(Au|f) = \frac{1}{2}|Au - f|^2$. Then the derivative of $\mathcal{E}$ with respect to u is*

$$\partial_u \mathcal{E}(Au|f) = A^*(Au - f). \tag{3.26}$$

Proof. Straightforward calculations lead to

$$\partial_{u_k}(Au)_m = \partial_{u_k} \sum_{n=1}^{N} A_{m,n}u_n = \sum_{n=1}^{N} A_{m,n}\delta_{m,k} = A_{m,k} = A^*_{k,m} \tag{3.27}$$

which in turn yields

$$\partial_{u_k}\left(\frac{1}{2}|Au - f|^2\right) = \partial_{u_k}\left(\frac{1}{2}\sum_{m=1}^{M}(Au - f)_m^2\right) \tag{3.28}$$

$$\overset{(3.27)}{=} \sum_{m=1}^{M} A^*_{k,m}(Au - f)_m = [A^*(Au - f)]_k. \tag{3.29}$$

■

Example 3.5.2 (MRI). In MRI, the forward operator can be decomposed as $B = \mathrm{Re}^* \circ F \circ Z^*$ with adjoint $B^* = Z \circ F^{-1} \circ \mathrm{Re}$. Therefore, the derivative of the data

fidelity with respect to the image is given by

$$\partial_v \mathcal{E}(Bv|g) = \frac{1}{\sigma^2} \operatorname{Re}(F^{-1} Z r) \tag{3.30}$$

with the residual $r \coloneqq Bv - g$, zero-filling Z, the inverse Fourier transform F^{-1} and the projection onto the real part Re. ▲

Example 3.5.3 (Parallel MRI). The derivative of the parallel MRI data fidelity (3.12) with B defined as (2.23) is given by

$$\partial_v \mathcal{E}(Bv|g) = \frac{1}{\sigma^2} \sum_{k=1}^{K} \operatorname{Re}\left(\overline{S_k} F^{-1} Z r_k\right) \tag{3.31}$$

with the residual for each receiver coil $r_k = B_k v - g_k$. ▲

Proof. Observe that for the parallel MRI operator the data fidelity becomes

$$\mathcal{E}(Bv|g) = \frac{1}{2\sigma^2} \sum_{k=1}^{K} |B_k v - g_k|^2 . \tag{3.32}$$

The claim follows from lemma 3.5.1 as $B_k^* = Z \circ F^{-1} \circ \overline{S_k} \circ \operatorname{Re}$. ■

We will now consider Poisson noise and compute the derivative of the PET data fidelity.

Lemma 3.5.4. *The derivative of the PET data fidelity* (3.16) *is given by*

$$\partial_u \mathcal{D}(Au + b|f) = A^*[\mathbb{1} - f \oslash (Au + b)], \tag{3.33}$$

where the symbol $\oslash$ *denotes point-wise division and* $\mathbb{1}$ *is a vector of ones with appropriate dimensions.*

Proof. Straight forward computations lead to

$$\partial_{u_k} \mathcal{D}(Au + b|f) = \partial_{u_k} \left\{ \sum_{m=1}^{M} \left[f_m \log\left(\frac{f_m}{(Au+b)_m} \right) + (Au + b - f)_m \right] \right\} \tag{3.34}$$

$$\overset{(3.27)}{=} \sum_{m=1}^{M} \left\{ -f_m A_{k,m}^* \frac{1}{(Au+b)_m} + A_{k,m}^* \right\} \tag{3.35}$$

$$= \sum_{m=1}^{M} \left\{ A_{k,m}^* \left[1 - \frac{f_m}{(Au+b)_m} \right] \right\} = \left\{ A^*[\mathbb{1} - f \oslash (Au+b)] \right\}_k . \tag{3.36}$$

■

Remark 3.5.5. The necessary condition for a minimum $\partial_u \mathcal{D}(Au+b|f) = 0$ multiplied point-wise by u, denoted by $\otimes$, and divided point-wise by $A^*\mathbb{1}$ yields a fixed point equation [23, 9]

$$u = u \oslash A^*\mathbb{1} \otimes A^*[f \oslash (Au + b)]. \tag{3.37}$$

This equation can be iteratively solved by the maximum likelihood expectation maximization algorithm (MLEM) [137]

$$u^{n+1} = u^n \oslash A^*\mathbb{1} \otimes A^*[f \oslash (Au^n + b)] \tag{3.38}$$

– the by far most common algorithm to reconstruct PET. One reason why this algorithm is so popular is the implicit positivity constraint which this iteration fulfils when the data and initial guess are positive. In addition, it is very easy to implement as it only requires a few lines of code with no extra parameters or line searches. On the downside, it is not easy to include arbitrary a priori information on the solution. Many extensions of this algorithm to include a priori information have some negative properties like the one-step-late algorithm [68] for which the regularization parameter cannot be chosen "too large" in order to be convergent and positivity preserving. Recently, techniques that split the likelihood from the prior have overcome this difficulty for total variation regularization [136]. ▲

3.6. Summary.

In this chapter we introduced a framework for reconstruction from a statistical point of view. Based on reasonable assumptions we derived the data fidelity terms for Gaussian and Poisson noise which are used for colour imaging and PET-MRI. Similarly, by expressing our beliefs of a multi-modality imaging system in form of a graphical model the objective functional often used for joint reconstruction can be rigorously derived from a statistical view point.

Chapter 4

Regularization based on Gradient Information

Regularization is a necessary ingredient for ill-posed inverse problems. On the one hand, it stabilizes the inversion so that even ill-conditioned problems become solvable. On the other hand, regularization provides a possibility to incorporate prior knowledge about the solution which is crucial if the solution is not unique. For many applications we do not know much about the values of the solution themselves but it is often reasonable to assume that the solution has similar values within a neighbourhood. Therefore, we consider regularization based on gradient (local neighbourhood) information.

4.1. Examples of Gradient based Regularization.

Example 4.1.1 (H^1-regularization). A classic regularization on the gradient is the L^2-norm of the gradient – equivalently the squared H^1-semi-norm of the image –

$$|u|_{\mathsf{H}^1}^2 \coloneqq \frac{1}{2}\int_\Omega |\nabla u(x)|^2 \,\mathrm{d}x. \tag{4.1}$$

This is a suitable regularization for imaging: It penalizes large gradients and therefore can be effectively utilized to remove noise. ▲

Example 4.1.2 (Total Variation). Another regularization on the gradient is *total variation* defined as

$$\mathbb{TV}(u) \coloneqq \int_\Omega |\nabla u(x)| \,\mathrm{d}x \tag{4.2}$$

for functions u that are smooth enough. It has been first studied in image processing in [120] for edge-preserved denoising. In contrast to H^1-regularization, which penalizes gradients that are twice as large four times as much, total variation penalizes these gradients "only" twice as much. Therefore, edges in an image – where the gradient is

very large – are penalized a lot less with total variation than with H^1-regularization. Later in this chapter, we will see why total variation is in fact edge-preserving while H^1-regularization is not. ▲

Example 4.1.3 (Joint Total Variation). Over the years there have been many attempts to generalize total variation to vector-valued (in particular colour) images. One of these extension to vector-valued images $w\colon \Omega \to \mathbb{R}^M$ is given by *joint total variation* [135, 74] – also called vectorial total variation –

$$\mathbb{TV}^{\mathcal{J}}(w) \coloneqq \int_\Omega \Big(\sum_{m=1}^{M} |\nabla w_m(x)|^2 \Big)^{1/2} \mathrm{d}x. \tag{4.3}$$

This is an *extension* in the sense that if all channels $m \neq m_0$ contain no structure, i.e. $\nabla w_m = 0$, it reduces to total variation in the m_0th channel. ▲

As we have seen in the examples above, many regularizations based on gradient information can be written in the general form

$$\mathcal{P}(u) = \int_\Omega f(\nabla u(x))\,\mathrm{d}x \quad \text{or short} \quad \mathcal{P}(u) = \int_\Omega f(\nabla u), \tag{4.4}$$

where the image is a function $u\colon \Omega \subset \mathbb{R}^D \to \mathbb{R}$, its gradient $\nabla u\colon \Omega \to \mathbb{R}^D$ and the *local regularization function* $f\colon \mathbb{R}^D \to \mathbb{R}$. For vector-valued images $w\colon \Omega \to \mathbb{R}^M$, we can fix all but one of the channels and so consider them separately.

4.2. Gâteaux Derivative.

The numerical results in this thesis will be obtained by smooth optimization, i.e. we will make use of derivatives to compute images with smaller and smaller function value. In a finite dimensional setting, we need the gradient of the objective function or equivalently, the partial derivatives with respect to all components of the image. In infinite dimensions – where we formulate our a priori knowledge – there is no such thing as partial derivatives and we need another notion of differentiability – the Gâteaux derivative [163, 162, 60, 81, 4]. We will calculate the Gâteaux derivative of gradient-based regularizations $\mathcal{P}$ in the following. As this notion is not very common and two different definitions of the Gâteaux derivative exist in the literature, we will state the one we will use here.

Definition 4.2.1 (Gâteaux derivative, [4]). *Let* X *be a Banach space and* $\mathcal{P}\colon \mathsf{X} \to \mathbb{R}$.

We call

$$D\mathcal{P}[u](v) = \lim_{\varepsilon \to 0^+} \frac{\mathcal{P}(u + \varepsilon v) - \mathcal{P}(u)}{\varepsilon} \tag{4.5}$$

the directional derivative *of $\mathcal{P}$ at u in the direction v provided the limit exists. Moreover, if there exists a bounded, linear operator $\tilde{u} \in \mathsf{L}(\mathsf{X}, \mathbb{R})$ such that $D\mathcal{P}[u](v) = \tilde{u}(v)$ for all $v \in \mathsf{X}$, we say that $\mathcal{P}$ is* Gâteaux differentiable *at u and we write $D\mathcal{P}[u] = \tilde{u}$.*

Assumption 4.2.2. *The domain $\Omega \subset \mathbb{R}^D$ needs to fulfil the assumptions of Green's first identity, cf. [48, p. 526, p. 534]. It is sufficient for us to think of a rectangular domain (or similar in arbitrary dimensions) such that an outer normal direction is defined almost everywhere on the boundary.*

For both the following theorem and for the numerical implementation we need that the images have vanishing normal derivatives on the boundary of the domain. This guarantees that there is no flow out of the image.

Definition 4.2.3. *The space of smooth images that can be extended to the boundary $\overline{\Omega} = \Omega \cup \partial\Omega$ can be defined as*

$$\mathsf{C}^1(\overline{\Omega}) \coloneqq \left\{ u \in \mathsf{C}^1(\Omega) \mid u \in \mathsf{C}(\overline{\Omega}), \nabla u \in \mathsf{C}(\overline{\Omega}, \mathbb{R}^N) \right\} \tag{4.6}$$

which has the sub-space of images with Neumann boundary conditions

$$\mathsf{C}^1_*(\overline{\Omega}) \coloneqq \left\{ u \in \mathsf{C}^1(\overline{\Omega}) \mid \partial_\nu u|_{\partial\Omega} = 0 \right\}, \tag{4.7}$$

where $\partial_\nu u$ is the derivative in the outer normal direction ν of the boundary $\partial\Omega$.

Remark 4.2.4. The space $\mathsf{C}^1_*(\overline{\Omega})$ is sufficient such that all formulae are well-defined. It is likely that this space as well as other assumptions stated here can be relaxed but an investigation in this direction is out of scope of this thesis.

In addition, when dealing with minimization problems, usually different spaces are chosen for different reasons. The most important one is to prove the existence of solutions. In this work, the continuous setting is a motivation for the discretized problem we solve and therefore, the focus has not been laid on questions of existence of solutions or more general well-posedness of the problem in the continuum. ▲

Theorem 4.2.5. *Let $f \in \mathsf{C}^2(\mathbb{R}^D)$ such that for any $u \in \mathsf{C}^1_*(\overline{\Omega})$ we have*

$$\langle \nabla f(\nabla u), \nu \rangle \,|_{\partial\Omega} = 0. \tag{4.8}$$

Then the Gâteaux derivative of $\mathcal{P}\colon \mathsf{C}^1_*(\overline{\Omega}) \to \mathbb{R}, \mathcal{P}(u) = \int_\Omega f(\nabla u)$ at any $u \in \mathsf{C}^1_*(\overline{\Omega})$ is

$$D\mathcal{P}[u] = -\operatorname{div}(\nabla f(\nabla u)). \tag{4.9}$$

For the proof of the theorem we need the following lemma.

Lemma 4.2.6. *Let $g\colon \mathbb{R}^D \to \mathbb{R}$ be differentiable and $x, y \in \mathbb{R}^D$. Then the derivative of $f\colon \mathbb{R} \to \mathbb{R}, f(\varepsilon) = g(x + \varepsilon y)$ with respect to ε at any point $\varepsilon_0 \in \mathbb{R}$ is given by*

$$f'(\varepsilon_0) = \langle \nabla g(x + \varepsilon_0 y), y \rangle \,. \tag{4.10}$$

Proof. Using $h\colon \mathbb{R} \to \mathbb{R}^D, h(\varepsilon) = x + \varepsilon y$ we can write f as $f(\varepsilon) = (g \circ h)(\varepsilon)$. Then we can apply the chain rule [47, p. 51] and obtain the stated result.

$$f'(\varepsilon_0) = \sum_{d=1}^{D} \partial_d g(h(\varepsilon_0)) \cdot (h_d)'(\varepsilon_0) = \sum_{d=1}^{D} [\nabla g(x + \varepsilon_0 y)]_d \cdot y_d = \langle \nabla g(x + \varepsilon_0 y), y \rangle \tag{4.11}$$

■

Proof of theorem 4.2.5. We approximate $\mathcal{P}(u+\varepsilon v)$ using Taylor's expansion around zero [47, p. 68, p. 70] for $g(\varepsilon) \coloneqq f(\nabla u + \varepsilon \nabla v)$ and lemma 4.2.6.

$$\varepsilon^{-1}\big[\mathcal{P}(u + \varepsilon v) - \mathcal{P}(u)\big] = \int_\Omega \varepsilon^{-1}\big(g(\varepsilon) - g(0)\big) = \int_\Omega \big(g'(0) + \mathcal{O}(\varepsilon)\big) \tag{4.12}$$

$$\overset{4.2.6}{=} \int_\Omega \langle \nabla f(\nabla u), \nabla v \rangle + \mathcal{O}(\varepsilon) \tag{4.13}$$

Using this approximation and Green's first identity [48, p. 526, p. 534] we derive the assertion.

$$D\mathcal{P}[u](v) = \lim_{\varepsilon \to 0} \varepsilon^{-1}\big[\mathcal{P}(u + \varepsilon v) - \mathcal{P}(u)\big] \overset{(4.13)}{=} \int_\Omega \langle \nabla f(\nabla u), \nabla v \rangle \tag{4.14}$$

$$= -\int_\Omega v \, \operatorname{div}(\nabla f(\nabla u)) + \int_{\partial\Omega} v \, \langle \nabla f(\nabla u), \nu \rangle \overset{(4.8)}{=} -\int_\Omega v \, \operatorname{div}(\nabla f(\nabla u)) \tag{4.15}$$

■

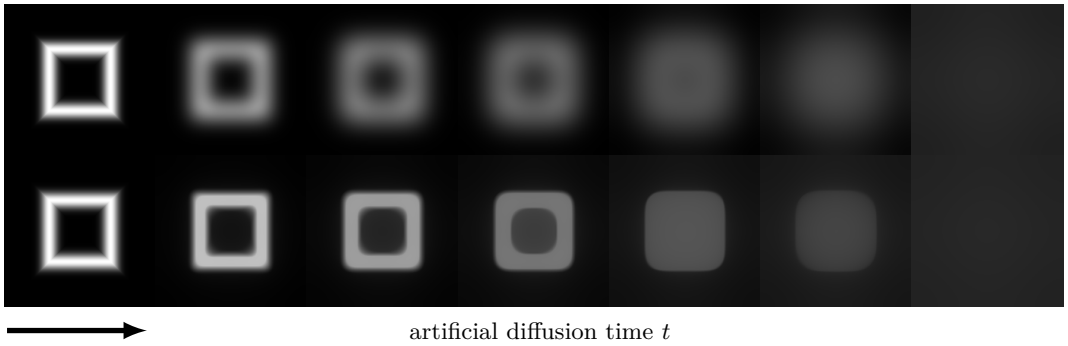


Figure 4.1. *Diffusive flow of* H^1*-regularization (top) and smooth total variation (bottom). From left to right are solutions of the diffusion equation* (4.18) *at increasing time points. The images on the right are the final state which are constant equal to the mean of the initial images.*

4.3. Diffusive Flow Analysis.

Many regularizations that act on gradient information have local regularization functions f such that

$$\nabla f(x) = \varkappa(x)x \tag{4.16}$$

with $\varkappa(x) \in \mathbb{R}^{D\times D}$. According to the theorem above, this results in a Gâteaux derivative of the form

$$D\mathcal{P}[u] = -\operatorname{div}\big(\varkappa(x,u)\nabla u\big). \tag{4.17}$$

Therefore, we can analyse a regularization $\mathcal{P}(u) = \int_\Omega f(\nabla u)$ not only in terms of its local regularization function f as in the beginning of this chapter but also in terms of its diffusivity $\varkappa$.

Minimizing a regularization functional which fulfils (4.16) with a gradient descent iteration corresponds to solving the partial differential equation

$$\partial_t u = -D\mathcal{P}[u] = \operatorname{div}\big(\varkappa\nabla u\big). \tag{4.18}$$

with an explicit Euler discretization. Therefore, these kind of regularizations are related to *diffusion / diffusive flow* [127, 1, 158, 146]. Two example diffusive flows for H^1-regularization and smooth total variation, cf. example 4.3.6, are shown in figure 4.1. It is evident that while H^1-regularization smooths the image everywhere, important edges are preserved much longer for smooth total variation.

The diffusivity that corresponds to a certain regularization can have very different

properties. In general, it is a matrix-valued function and depends on both the location and the underlying image.

Definition 4.3.1 (Characterisation of Scalar-Valued Diffusion, [158]). *The diffusive flow that corresponds to a diffusivity $\varkappa(x,u) \in \mathbb{R}^{D\times D}$ is called*

1. homogeneous*, if the diffusivity does not depend on the spatial location, i.e. $\varkappa(x,u) = \varkappa$, otherwise* inhomogeneous.

2. linear*, if the diffusivity does not depend on the underlying image, i.e. $\varkappa(x,u) = \varkappa(x)$, otherwise* non-linear.

3. isotropic*, if the diffusivity is a multiple of the identity matrix I, i.e. $\varkappa(x,u) = \kappa(x,u)I$ with $\kappa(x,u) \in \mathbb{R}$, otherwise* anisotropic.

Remark 4.3.2. The diffusion of Perona and Malik [127] is sometimes referred to as anisotropic but it is isotropic and non-linear in our terminology. ▲

Remark 4.3.3. When analysing the diffusivity $\varkappa$ one needs to be very careful as one can see from (4.18) that any matrix with ∇u in its kernel can be added and results in the same diffusion process [146]. As an example, let $\nabla u \neq 0$, $n := \nabla u / |\nabla u|$ and choose a normalized vector $t \in \mathbb{R}^N$ orthogonal to ∇u, i.e. $\langle t, \nabla u\rangle = 0$. Then the diffusivities $\varkappa_1 = I$ and $\varkappa_2 = R\,\mathrm{diag}([1;s])R^T$ for any $s \in \mathbb{R}$ and $R = [n,t]$ generate the same diffusion as $\varkappa_1\nabla u = \varkappa_2\nabla u$. Note, that all vectors are treated as column vectors and we have used MATLAB notation to define $\varkappa_2$ and R. ▲

Example 4.3.4 (H^1-Regularization). H^1-regularization fits into the general framework $\int_\Omega f(\nabla u)$ with the local regularization function $f(x) = \frac{1}{2}|x|^2$. As the gradient of f is $\nabla f(x) = x$, its Gâteaux derivative is given by

$$D\,|\cdot|_{\mathsf{H}^1}[u] = -\operatorname{div}(\kappa\nabla u) = -\Delta u \tag{4.19}$$

with the diffusivity $\kappa = 1$ being constant. Therefore, this corresponds to *linear*, *homogeneous* and *isotropic* diffusion. The corresponding diffusion equation describes the flow of heat within an isotropic medium. The steady state ($t \to \infty$) of this equation is constant where the constant is equal to the mean-value of the initial image. This is illustrated in figure 4.1 as well. ▲

Example 4.3.5 (Total Variation). For total variation the local regularization function $f(x) = |x|$ is not differentiable in zero. Everywhere else it is and the gradient there is

$\nabla f(x) = x/|x|$. Therefore one might write the Gâteaux derivative as

$$D\mathbb{TV}[u] = -\operatorname{div}\left(\frac{1}{|\nabla u|}\nabla u\right) \tag{4.20}$$

keeping in mind that it is well-defined only for $|\nabla u| \neq 0$. This regularization leads to *inhomogeneous* diffusion as the diffusivity $\kappa = 1/|\nabla u|$ is spatially variant. In addition, as the diffusivity is scalar and it depends on the image itself the diffusion is *isotropic* and *non-linear*.

Formally, this diffusion scheme is edge-preserving as $\kappa \to 0$ when $|\nabla u| \to \infty$ so that for steeper and steeper edges the diffusivity tends to zero. Furthermore, it favours piecewise constant regions within these edges as $\kappa \to \infty$ when $|\nabla u| \to 0$. Therefore, solutions that are obtained with total variation regularization tend to have a comic-like appearance which makes them ideal for segmentation. ▲

Example 4.3.6 (Smooth Total Variation). The local regularization function for total variation is not differentiable in zero. This is not only a theoretical problem but leads to division by zero in flat areas. There are many ways to *smooth* the total variation functional. One option is to use a smooth approximation of the Euclidean norm $|x|^2 \approx |x|_\beta^2 \coloneqq |x|^2 + \beta^2$ with $\beta > 0$. This smoothed norm is differentiable for all x and its gradient is $x/|x|_\beta$, which is well-defined for all x as $|x|_\beta \geq |0|_\beta = \beta > 0$. Then the Gâteaux derivative of the smoothed total variation functional is

$$D\mathbb{TV}_\beta[u] = -\operatorname{div}\left(\frac{1}{|\nabla u|_\beta}\nabla u\right). \tag{4.21}$$

Here the diffusivity $\kappa = 1/|\nabla u|_\beta$ approximates the "real" total variation diffusivity well in the sense that for $|\nabla u| \to 0$ the diffusivity tends to $1/\beta$. With better and better approximations, i.e. $\beta \to 0$, this reflects the behaviour of the diffusivity of the usual total variation. ▲

4.3.1. Vector-Valued Images. In case of a vector-valued image $w\colon \Omega \to \mathbb{R}^M$ with $M > 1$ the corresponding Gâteaux derivative can often be written as

$$D\mathcal{P}[w_m] = -\operatorname{div}(\varkappa_m(x, w)\nabla w_m), \quad m = 1, \ldots, M \tag{4.22}$$

and we extend the definitions 4.3.1 to the vector-valued case if they hold for all channels.

Definition 4.3.7 (Characterisation of Vector-Valued Diffusion). *Vector-valued diffusion*

is called channel-by-channel *if the diffusivity for each channel only depends on the very same channel, i.e.*

$$\varkappa_m(x, w) = \varkappa_m(x, w_m), \quad m = 1, \dots, M, \tag{4.23}$$

otherwise, we will call it channel-coupled.

Example 4.3.8 ((Smooth) Separate Total Variation). The simplest way to generalize the smooth total variation to vector-valued images $w\colon \Omega \to \mathbb{R}^M$ is given by

$$\mathbb{TV}_\beta^{\mathcal{S}}(w) = \sum_{m=1}^{M} \mathbb{TV}_\beta(w_m) = \int_\Omega \sum_{m=1}^{M} |\nabla w_m|_\beta \, . \tag{4.24}$$

Its Gâteaux derivative with respect to any channel w_m is

$$D\mathbb{TV}_\beta^{\mathcal{S}}[w_m] = -\operatorname{div}\left(\frac{1}{|\nabla w_m|_\beta}\nabla w_m\right), \quad m = 1, \dots, M \tag{4.25}$$

so that the generated diffusion is *channel-by-channel* total variation diffusion. While it is very simple and easy to implement for arbitrary number of channels, it does not exploit the coupling seen often in multi-modality images. ▲

Example 4.3.9 ((Smooth) Colour Total Variation). One of the first generalizations of total variation to vector-valued images (in particular colour images) that couple information from the different channels was proposed by [15],

$$\mathbb{TV}_\beta^{\mathcal{C}}(w) = \left[\sum_{m=1}^{M} \big(\mathbb{TV}_\beta(w_m)\big)^2\right]^{1/2} = \left[\sum_{m=1}^{M} \Big(\int_\Omega |\nabla w_m|_\beta\Big)^2\right]^{1/2}, \tag{4.26}$$

and is often referred to as *colour total variation*. It couples the channels of a vector-valued image globally – the total variation of each channels is first computed and then combined. This property is also referred to as *global pooling* [80] and the opposite of joint total variation, cf. the beginning of this chapter or the next example, where the information is coupled locally.

Although it does not fit directly into the framework of section 4.2, its Gâteaux derivatives can be easily computed as

$$D\mathbb{TV}_\beta^{\mathcal{C}}[w_m] = -\frac{\mathbb{TV}_\beta(w_m)}{\mathbb{TV}_\beta^{\mathcal{C}}(w)} \operatorname{div}\left(\frac{1}{|\nabla w_m|_\beta}\nabla w_m\right), \quad m = 1, \dots, M. \tag{4.27}$$

The generated diffusion is *channel-coupled* but again we see that the coupling is *global*.

It turns out that this coupling is weak and cannot fully exploit the correlations between colour channels as we have seen in the introduction. ▲

Example 4.3.10 ((Smooth) Joint Total Variation). Similarly to total variation, we can define a "smooth" version of joint total variation, cf. (4.3), as

$$\mathbb{TV}_\beta^{\mathcal{J}}(w) = \int_\Omega \left(\sum_{m=1}^{M} |\nabla w_m|^2 + \beta^2 \right)^{1/2} \tag{4.28}$$

with Gâteaux derivative

$$D\mathbb{TV}_\beta^{\mathcal{J}}[w_m] = -\operatorname{div}\left(\frac{1}{\left(\sum_{k=1}^{M} |\nabla w_k|^2 + \beta^2\right)^{1/2}} \nabla w_m \right), \quad m = 1, \dots, M. \tag{4.29}$$

Joint total variation results in *channel-coupled*, *non-linear*, *inhomogeneous* but *isotropic* diffusion. Moreover, the diffusivity is the same for all channels. As the channels are coupled locally this is sometimes referred to as *local pooling* [80].

Whereas total variation promotes images that have a sparse support of the gradient, joint total variation promotes images with gradients of joint support. This can be seen in two ways. First, from the functional formulation (4.28), we see that if there is a large gradient in one of the channels, the gradients in the other channels are less penalized. Second, it can also be seen from the diffusivity: a large gradient in one of the channels reduces the diffusivity in all channels close to zero so that there is little diffusion in any of the other channels. ▲

4.4. Summary.

The (Gâteaux-) derivative of functionals that are based on gradients is often linked to diffusion. We have seen several examples – more will follow in the next chapter – where the derivative is of the form $-\operatorname{div}(\varkappa \nabla u)$ so that we can analyse these functionals also in terms of their diffusivity $\varkappa$. These diffusivities can be characterized as linear/non-linear, homogeneous/inhomogeneous, isotropic/anisotropic and for vector-valued images u channel-by-channel/channel-coupled. All diffusivities of the functionals considered so far were isotropic, but this will change in the next chapter.

Chapter 5

Parallel Level Sets

As we have seen in the introduction, multi-modality images can show similar structures in the different channels. In this chapter, we will have a closer look at the structure of scalar- and vector-valued images. We will develop a mathematical notion how similar structures can be modelled, called *parallel level sets*. Based on this, we will propose several priors that all promote similar images according to this notion. We analyse some of their properties and compare them to related priors in the literature.

5.1. Measuring Common Structure by Parallel Level Sets.

Before we can measure common structure we need to specify what we mean by *structure of an image*.

5.1.1. The Structure of an Image.

Scalar Valued Images. An image – modelled as a function $u\colon \Omega \subset \mathbb{R}^D \to \mathbb{R}$ – maps a spatial location $x \in \Omega$ onto a "colour" $u(x) \in \mathbb{R}$. It is a somewhat philosophical question what the structure of an image is though everyone has an intuitive opinion. It is intuitively clear that structure $\mathcal{S}$ should be invariant under global "colour" transformations

$$\text{colour invariance} \qquad \mathcal{S}(u) = \mathcal{S}(f(u)) \tag{5.1}$$

with $f\colon \mathbb{R} \to \mathbb{R}$ injective, i.e. it is not the actual colours that matter. "Injective" means that not two different colours are mapped to the same colour. If this was the case, then structure might disappear.

A model for structure that fulfils this assumption is the *absolute orientation of the gradient*, i.e. the orientation modulo a change of the sign. It is well known that for differentiable functions the gradient is orthogonal to the level sets at every point [47].

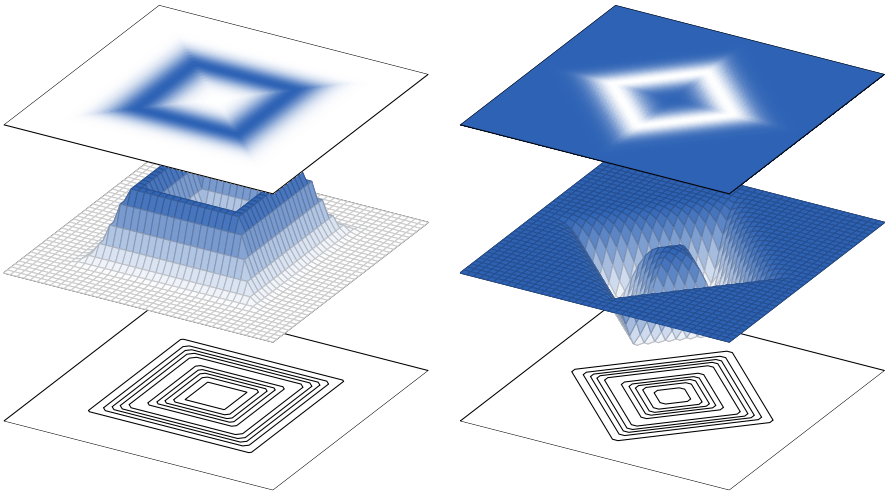


Figure 5.1. *Images can be represented as a 2D plot with colours/grey scales (top) or as a surface plot (center) where the height is given by the image value at a certain location. Level sets (bottom) capture the structure of images – shown here for a few values.*

Therefore, one could say that the structure of an image is contained in its level sets – a formal definition is given below. This ad hoc statement agrees with [31] where it has been argued that the structure of natural images is contained in its level sets.

Definition 5.1.1 (Level Sets). *Let $u\colon \Omega \to \mathbb{R}$ be a function. For each $s \in \mathbb{R}$ the set*

$$u^{-1}(s) \coloneqq \{x \in \Omega \mid u(x) = s\} \tag{5.2}$$

is called the s-level set *or short* level set *of the function u.*

Two example images in two dimensions are depicted in figure 5.1 as a two dimensional "colour" image and as a surface plot. Their level sets which encode the structure of these images are shown at the bottom.

5.1.2. Vector-Valued Images. Vector-valued images $w\colon \Omega \subset \mathbb{R}^D \to \mathbb{R}^M$ arise in several applications such as colour imaging and multi-modality medical imaging. As we have seen in the introduction, these images exhibit similar structures, as they are taken from the same scenery or anatomy.

It has been argued that such images essentially decompose like

$$w_m(x) = \tau_m(x)\rho(x), \tag{5.3}$$

where ρ describes the structure and τ is a material property [91, 80]. As material does not change arbitrarily, it is reasonable to assume that τ_m is slowly varying or piecewise

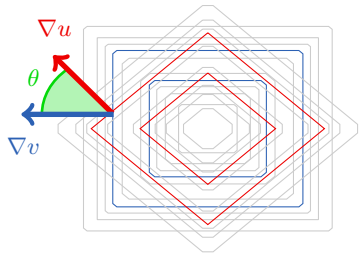


Figure 5.2. *The two test images from figure 5.1 and some of their level sets. As gradients are perpendicular to level sets, structural similarity can be measured in terms of the angle between the gradients.*

constant. In colour imaging, the Lambertian image model yields (5.3) where $\tau \in [0, 1]$ is the albedo that depends on the frequency of the incoming light and ρ the inner product between the surface normal and the direction of incoming light [91]. Similarly, in medical images ρ describes the anatomy and τ tissue properties like proton density, T1, T2 in MRI or tracer concentration in PET. Such a model has also been used in dual-energy CT reconstructions [62]. Under these assumptions it is reasonable to assume that the colour channels in an RGB image and PET-MRI images have the same structure and can be measured with image gradients and level sets. We will discuss this assumption further in chapter 10.

In our model the structure of an image is related to the absolute orientation of its gradients. Therefore, to establish a measure on the structural difference of two images we shall make use of parallel vectors.

5.1.3. Measuring Parallel Vectors.

Definition 5.1.2 (Parallel Vectors). *Two vectors $x, y \in \mathbb{R}^D$ are called* parallel, *in symbols $x \parallel y$, if there exists a $s \in \mathbb{R}$ such that either $x = sy$ or $y = sx$.*

For two vectors $x, y \in \mathbb{R}^D$ and the angle between them $\theta \in [0, 2\pi)$, it holds

$$\langle x, y \rangle = \cos(\theta) \, |x| \, |y| \,, \tag{5.4}$$

which follows from the law of cosines. Based on this equation, we can derive a "measure" of parallelism

$$d(x, y) := |x| \, |y| - |\langle x, y \rangle| \,. \tag{5.5}$$

Some of its properties are summarized in the following proposition.

Proposition 5.1.3. *For all $x, y \in \mathbb{R}^D$, there is*

$$\begin{aligned} &\textit{1) symmetry:} && d(x,y) = d(y,x) && (5.6)\\ &\textit{2) non-negativity:} && d(x,y) \geq 0 && (5.7)\\ &\textit{3) identity of indiscernibles:} && d(x,y) = 0 \Leftrightarrow x \parallel y. && (5.8) \end{aligned}$$

Proof. All the properties follow easily from (5.4) as both multiplication and the scalar product are commutative and $1 - |\cos(\theta)| \geq 0$. In addition, for non-zero vectors there is $x \parallel y \Leftrightarrow \theta = 0$ or $\theta = \pi$. ∎

Remark 5.1.4. This distance is not a metric as it does not fulfil the triangle inequality: For non-parallel vectors x and y there is

$$0 < d(x,y) \not\leq d(x,0) + d(0,y) = 0. \tag{5.9}$$

With these properties it can be either seen as a "pre-metric" or a "semi-metric" on the equivalence classes of parallel vectors. ▲

Remark 5.1.5. All these statements hold in infinite dimensions as well if one defines the angle between two infinite dimensional vectors to fulfil (5.4). But as the vectors considered here will be gradients of functions $u\colon \Omega \subset \mathbb{R}^D \to \mathbb{R}$, these are intrinsically of finite dimension. ▲

All properties of d depend on inequalities so that we can generalize it with strictly increasing functions $\varphi, \psi\colon [0,\infty) \to [0,\infty), \varphi(0) = 0$ and measure parallelism with

$$f_{\mathrm{PL}}(x,y) \coloneqq \varphi\big[\psi(|x|\,|y|) - \psi(|\langle x, y\rangle|)\big]. \tag{5.10}$$

5.1.4. Measuring Common Structure of Images.

Definition 5.1.6 (Parallel Level Sets). *Let $u, v \in \mathsf{C}^1(\Omega)$. We say u and v have* parallel level sets *denoted by $u \parallel v$ if their gradients are parallel at every point, i.e. for all $x \in \Omega$ there is $\nabla u(x) \parallel \nabla v(x)$.*

Remark 5.1.7. Every image has parallel level sets with itself, i.e. $u \parallel u$. As every vector is parallel to the zero vector, every image has parallel level sets with constant images. ▲

Definition 5.1.8 (Measure of Parallel Level Sets). *Let $\varphi, \psi\colon [0,\infty) \to [0,\infty), \varphi(0) = 0$ be strictly increasing functions. We define the generalized measure of parallel level sets as* $\mathrm{PL}\colon \mathsf{C}^1(\Omega) \times \mathsf{C}^1(\Omega) \to \mathbb{R}$ *with*

$$\mathrm{PL}(u,v) \coloneqq \int_\Omega f_{\mathrm{PL}}(\nabla u(x), \nabla v(x))\,\mathrm{d}x \tag{5.11}$$

$$\overset{(5.10)}{=} \int_\Omega \varphi[\psi(|\nabla u(x)|\,|\nabla v(x)|) - \psi(|\langle \nabla u(x), \nabla v(x)\rangle|)]\,\mathrm{d}x \tag{5.12}$$

$$= \int_\Omega \varphi[\psi(|\nabla u|\,|\nabla v|) - \psi(|\langle \nabla u, \nabla v\rangle|)]. \tag{5.13}$$

Proposition 5.1.9. *For all $u, v \in \mathsf{C}^1(\Omega)$ there is*

1) symmetry:	$\mathrm{PL}(u,v) = \mathrm{PL}(v,u)$	(5.14)
2) non-negativity:	$\mathrm{PL}(u,v) \geq 0$	(5.15)
3) identity of indiscernibles:	$\mathrm{PL}(u,v) = 0 \Leftrightarrow u \parallel v.$	(5.16)

Proof. The assertions follow directly from properties of the integral and proposition 5.1.3. ■

Remark 5.1.10. A precise statement of 3) in proposition 5.1.9 must include some information on the measure like "almost everywhere" for we neglect it here for simplicity. ▲

Measuring structure in terms of parallel level sets fulfils our initial requirement that colour transformations do not affect structure.

Proposition 5.1.11 (Parallel Level Sets and Colour Transformations). *Let $f \in \mathsf{C}^1(\mathbb{R})$ be injective and $u, v \in \mathsf{C}^1(\Omega)$. Then $u \parallel v$ if and only if $u \parallel (v \circ f)$.*

Proof. For every x there is $\nabla(v \circ f)(x) = f'(x)\nabla v(x)$. From the injectivity, it follows that $f'(x) \neq 0$ and therefore the claim. ■

5.2. Different Concepts based on Parallel Level Sets.

5.2.1. (Symmetric) Parallel Level Sets Prior. By definition, every vector is parallel to the null vector so that if one of the images is flat, then the other image is not regularized. We therefore propose to alter the local regularization function of parallel level sets.

$$f_{\mathrm{PL}_\beta}(x,y) \coloneqq \varphi[\psi(|x|_\beta\,|y|_\beta) - \psi(|\langle x, y\rangle|_{\beta^2})] \tag{5.17}$$

Here we use the same notation $|x|_\beta^2 := |x|^2 + \beta^2$ as introduced for "smooth" total variation, cf. example 4.3.6.

Remark 5.2.1. In general, we could define the smoothed norms in other ways while still having the same properties. Let

$$\mathsf{E} := \left\{ \eta \in \mathsf{C}^1([0,\infty)) \mid \eta(0) > 0, \eta \geq id, \lim_{s\to\infty} \eta(s) - s = 0, \lim_{s\to 0} \eta'(s)/s > 0 \right\} \tag{5.18}$$

and define for $\eta \in \mathsf{E}$ the smoothed norm $|x|_\eta := \eta(|x|)$. Note that $|\cdot|_\beta \in \mathsf{E}$. We will use the notation $x_\eta := x/|x|_\eta$ which is well defined for all $x \in \mathbb{R}^D$. These smoothed norms have the following properties.

1) smoothness: $$|\cdot|_\eta \in \mathsf{C}^1(\mathbb{R}^D) \tag{5.19}$$

2) normalization: $$|x|_\eta \leq 1, \quad \text{for all } x \in \mathbb{R}^D \tag{5.20}$$

3) approximation: $$\lim_{|x|\to\infty} |x|_\eta - |x| = 0 \tag{5.21}$$

Let $\eta_1, \eta_2, \eta_3 \in \mathsf{E}$. A generalization of (5.17) is

$$\varphi[\psi(|x|_{\eta_1} |y|_{\eta_2}) - \psi(|\langle x, y\rangle|_{\eta_3})], \tag{5.22}$$

but we are not going to pursue this any further in this thesis. ▲

Let us now focus on some special cases for φ and ψ.

Example 5.2.2 (Linear Parallel Level Sets). We call the special case $\varphi, \psi = id$, i.e.

$$f_{\mathrm{PL}_\beta^{\mathcal{L}}}(x,y) := |x|_\beta |y|_\beta - |\langle x, y\rangle|_{\beta^2}\,, \tag{5.23}$$

the *linear parallel level sets* prior. It is a generalization of total variation in the sense that if no structure is present in all but one channel, it reduces to smooth total variation in this channel. ▲

Example 5.2.3 (Quadratic Parallel Level Sets). The prior that corresponds to the choice $\varphi(s) = (\beta^4 + s)^{1/2} - \beta^2, \psi(s) = s^2$, i.e.

$$f_{\mathrm{PL}_\beta^{\mathcal{Q}}}(x,y) := \left(\beta^4 + |x|_\beta^2 |y|_\beta^2 - |\langle x, y\rangle|_{\beta^2}^2\right)^{1/2} - \beta^2 \tag{5.24}$$

$$= \left[\beta^4 + \beta^2(|x|^2 + |y|^2) + |x|^2 |y|^2 - |\langle x, y\rangle|^2\right]^{1/2} - \beta^2, \tag{5.25}$$

will be called the *quadratic parallel level set* prior.

The quadratic parallel level set prior is not only an extension of total variation but also of joint total variation. It is an extension of joint total variation in the sense that if the channels are aligned, i.e. $|\nabla u|^2 |\nabla v|^2 - \langle \nabla u, \nabla v \rangle^2 = 0$, it reduces to smooth joint total variation.

This functional has been proposed for colour image processing with a motivation from Riemannian geometry [139]. ▲

Remark 5.2.4. φ and ψ are chosen to "invert" each other so that the prior reduces to "smooth" total variation if one of the images is flat. ▲

Remark 5.2.5. We will see in chapter 6 that critical properties of the prior are due to ψ, hence, they are named after the choice of ψ. ▲

Example 5.2.6 (Cross-Gradients). Another special case is the *cross-gradients* functional [63] – as it involves the cross-product of two gradients –

$$\frac{1}{2} \int_\Omega |\nabla u \times \nabla v|^2 = \frac{1}{2} \int_\Omega |\nabla u|^2 |\nabla v|^2 - \langle \nabla u, \nabla v \rangle^2 , \tag{5.26}$$

which has been used for joint reconstruction in geophysics [63, 64, 65, 74] and as a one-sided prior in spectral imaging [7, 110]. This is a special case of (5.17) with $\varphi = id, \psi(s) = \frac{1}{2}s^2$ and $\beta = 0$. ▲

Remark 5.2.7. Following the same arguments as for the general parallel level sets functional, we can define a family of *normalized* parallel level sets functionals

$$\int_\Omega \varphi[\psi(1) - \psi(|\langle (\nabla u)_\eta, (\nabla v)_\eta \rangle|)] \tag{5.27}$$

for $\eta \in \mathsf{E}$ and using the notation introduced in remark 5.2.1. It has been been applied to multi-modality image registration [75] and face recognition [147] for the choice $\varphi = id, \psi(s) = s^2$ and $\eta = |\cdot|_\beta$. While it is beneficial for a similarity measure to be independent of the scale of the two images, for image reconstruction the scale of the gradients has to be penalized to suppress noise. ▲

Remark 5.2.8 (Extensions to Higher Dimensions). A way to generalize the parallel level set notion to arbitrary dimensions $w \colon \Omega \to \mathbb{R}^M$ is a *pairwise* approach. For any

finite number of channels M, we define the cost functional

$$\int_\Omega \varphi\left[\sum_{m=1}^{M}\sum_{k=m+1}^{M} \psi(|\nabla w_m|_\beta\,|\nabla w_k|_\beta) - \psi(|\langle\nabla w_m, \nabla w_k\rangle|_{\beta^2})\right]. \tag{5.28}$$

For $\beta = 0$, this is a generalization in the sense that if all but two channels are flat, it reduces to the bi-modal version we discussed before. ▲

5.2.2. Asymmetric Parallel Level Sets. The previous framework based on parallel level sets has been designed in a symmetric fashion which is well suited if the modalities have similar properties and therefore can (or maybe even should) be treated similarly. However, sometimes the imaging modalities have very different properties but still structure from another modality can be used. Therefore, we define the *asymmetric* parallel level sets prior

$$f_{\mathrm{APL}_\beta}(x;y) \coloneqq \varphi[\psi(|x|_\beta) - \psi(|\langle x, y_\eta\rangle|_\beta)] \tag{5.29}$$

with strictly monotonically increasing functions ψ, φ and $y_\eta \coloneqq y/\eta(|y|)$ with $\eta \in \mathsf{E}$, e.g. $\eta = |\cdot|_\gamma$, cf. example 5.2.1.

Similar to symmetric parallel level sets, we can define a linear and quadratic version.

Example 5.2.9 (Linear Asymmetric Parallel Level Sets). The special case $\varphi, \psi = id$

$$f_{\mathrm{APL}^{\mathcal{L}}_\beta}(x;y) \coloneqq |x|_\beta - |\langle x, y_\eta\rangle|_\beta \tag{5.30}$$

is called *linear asymmetric parallel level sets*. For $y = 0$, it reduces to smooth total variation. ▲

Example 5.2.10 (Quadratic Asymmetric Parallel Level Sets). The special case

$$f_{\mathrm{APL}^{\mathcal{Q}}_\beta}(x;y) \coloneqq \left(\beta^2 + |x|^2 - \langle x, y_\eta\rangle^2\right)^{1/2} \tag{5.31}$$

that corresponds to $\varphi(s) = (\beta^2 + s)^{1/2}, \psi(s) = s^2$ and is called *quadratic asymmetric parallel level sets*. If $y \neq 0$, we define the orthogonal component of x with respect to y as $x^\perp \coloneqq x - \langle x, y/\,|y|\rangle\, y$. It turns out that

$$f_{\mathrm{APL}^{\mathcal{Q}}_\beta}(x;y) = |x^\perp|_\beta \tag{5.32}$$

so that quadratic asymmetric parallel level sets corresponds to total variation on the orthogonal component. ▲

Example 5.2.11 (Kaipio). It was proposed in [86] to use the regularization

$$\frac{1}{2}\int_\Omega |A(x)\nabla u(x)|^2 \,\mathrm{d}x \tag{5.33}$$

where $A(x)$ is symmetric and positive-semidefinite, and related to an a priori defined and smoothly varying field $v\colon \Omega \to \mathbb{R}$ such that

1. If $\nabla u(x) \parallel \nabla v(x)$, then there is $A(x)\nabla u(x) = \rho(|\nabla v(x)|)\nabla u(x)$ with ρ being an edge indicator function, i.e. $\rho(0) = 1$ and $\lim_{s\to\infty} \rho(s) = 0$.
2. If $\nabla u(x) \perp \nabla v(x)$, then there is $A(x)\nabla u(x) = \nabla u(x)$.

In general there might be other options but the only option they presented is $\rho(s) \coloneqq [\tilde{\eta}(s)]^{-2}, \tilde{\eta}(s) \coloneqq (s^2+1)^{1/2}$ and

$$A \coloneqq I - (\nabla v)_{\tilde{\eta}}(\nabla v)_{\tilde{\eta}}^T. \tag{5.34}$$

▲

5.3. Linear versus Quadratic Parallel Level Sets.

Although the framework so far was generic, we will see in the following that the actual behaviour of the (symmetric) parallel level sets functional as a prior will depend on the actual choice of φ and ψ. Parts of this holds for the asymmetric case as well, but we leave a complete study of the asymmetric case for further work.

Example 5.3.1 (Parallel Level Sets as a Function of One Vector). First, we have a look at the local regularization functions of linear (5.23) and quadratic parallel level sets (5.24) as a function of one vector with the other one being set to $y = [1; -1]$. These functions for a two dimensional domain are plotted in figure 5.4. In the case of $\beta = 0$, both functions are constant zero for vectors parallel to y. While linear parallel level sets allows for a larger deviation for larger vectors, quadratic parallel level sets penalizes the norm of the orthogonal component to y independent of the total length. For larger β and for both cases, the level sets are smoothed and become more circular around zero. ▲

Example 5.3.2 (Rotation and Parallel Level Sets). In this example, we test the two parallel level sets functionals on rotation. Therefore, we take a vector $x \in \mathbb{R}^D$ and for

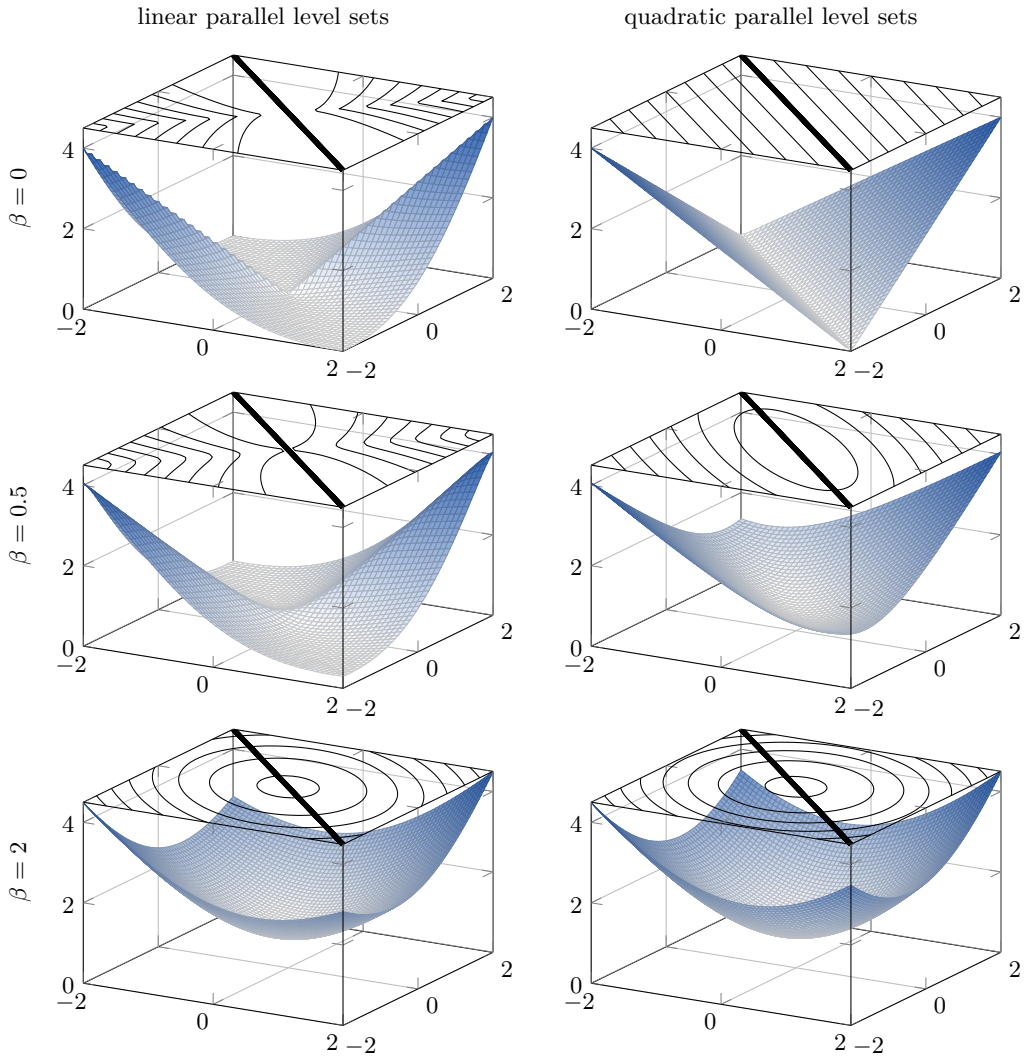


Figure 5.3. *The local regularization function of linear* (5.23) *and quadratic parallel level sets* (5.24) *as a function of one vector with the other one being set to* $y = [1; -1]$*. On top are some of their level sets as well as the preferred direction along* y*. The parameter* β *changes the shape of the function locally around zero.*

any angle $\theta \in [0, 2\pi)$ we define a rotated version of x which we will denote by $x(\theta)$. The rotational axis does not matter for this experiment. The vector $x(\theta)$ has the same length $|x(\theta)| = |x|$ and the inner product satisfies $\langle x, x(\theta)\rangle = \cos(\theta)\,|x|^2$. Then the local regularization functions of linear and quadratic parallel level sets (as a function of θ instead of x and y) are

$$f_{\mathrm{PL}_\beta^{\mathcal{L}}}(\theta) = \left(|x|^2 + \beta^2\right)^{1/2}\left(|x|^2 + \beta^2\right)^{1/2} - \left(\cos^2(\theta)\,|x|^4 + \beta^4\right)^{1/2} \tag{5.35}$$

$$= |x|^2 + \beta^2 - \left(\cos^2(\theta)\,|x|^4 + \beta^4\right)^{1/2} \tag{5.36}$$

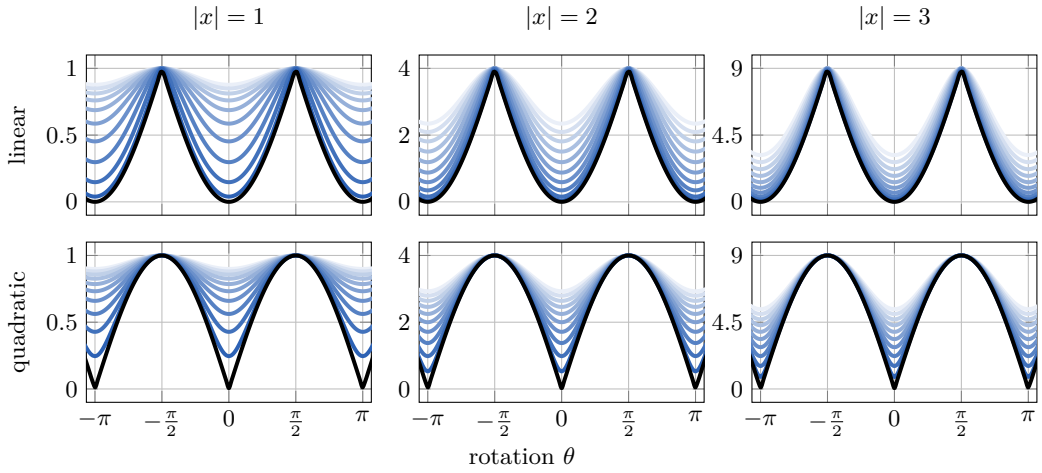


Figure 5.4. *Effect of rotation on the parallel level sets functionals. The regularization functions for linear parallel level sets* (5.36) *(top row) and quadratic level sets* (5.38) *(bottom row) as a function the rotation angle θ. These are plotted for $|x| = 1, 2, 3$ with $\beta = 0$ in black and decreasing colour intensity for $\beta = 0.2, 0.4, \ldots, 2$. See the text for more information.*

and

$$f_{\mathrm{PL}^2_\beta}(\theta) = \left[\beta^4 + (|x|^2 + \beta^2)(|x|^2 + \beta^2) - (\cos^2(\theta)\,|x|^4 + \beta^4)\right]^{1/2} - \beta^2 \tag{5.37}$$

$$= \left(\beta^4 + 2\beta^2\,|x|^2 + \sin^2(\theta)\,|x|^4\right)^{1/2} - \beta^2. \tag{5.38}$$

These functions are plotted in figure 5.4. Note first, that they essentially scale quadratically in $|x|$ and they both vanish for $|x| = 0$ (not shown). In addition, the behaviour of these functions for $\beta = 0$ – shown in black – is very different. While linear parallel level sets, i.e. $|x|^2\,(1 - |\cos(\theta)|)$, has a broad valley at $\theta = 0 \pm \pi$, quadratic parallel level sets, i.e. $|x|^2\,|\sin(\theta)|$, has a narrow valley. This means that a small variation in rotation from the optimal value is less penalized by the linear parallel level sets functional than by the quadratic one. For larger β or equivalently decreasing colour intensity, both functions tend to penalize deviating vectors ($|\theta| > 0$) less and less. In addition, for any fixed $\beta > 0$, the alignment becomes more important as $|x|$ grows. ▲

Example 5.3.3 (Scaling and Parallel Level Sets). This example is concerned with the effect of scaling on the parallel level sets functionals. As the effect is different whether we scale only one vector, both vectors simultaneously or both vectors independently; we consider all three cases separately.

Let $x, y \in \mathbb{R}^D$ be vectors of length s, t and θ the angle between them. With these definitions, consider the local regularization functions for parallel level sets as functions

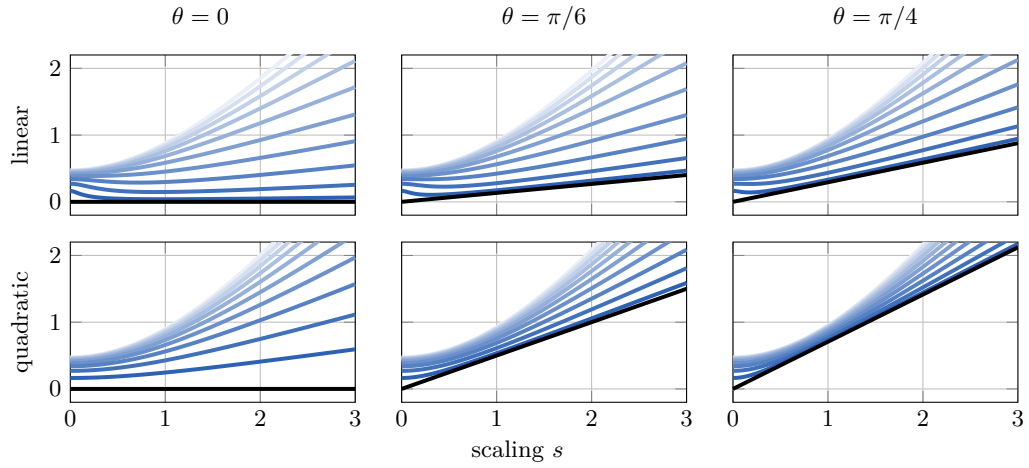


Figure 5.5. *Effect of scaling* one *vector on the parallel level sets functionals. The regularization functions for linear parallel level sets* (5.39) *(top row) and quadratic parallel level sets* (5.40) *(bottom row) as a function of the scaling s of one image. These are plotted for $t = 1$ and $\theta = 0, \pi/6, \pi/4$. In each axes the graphs are plotted for $\beta = 0$ in black and $\beta = 0.2, 0.4, \ldots, 2$ with decreasing intensity. See the text for more information.*

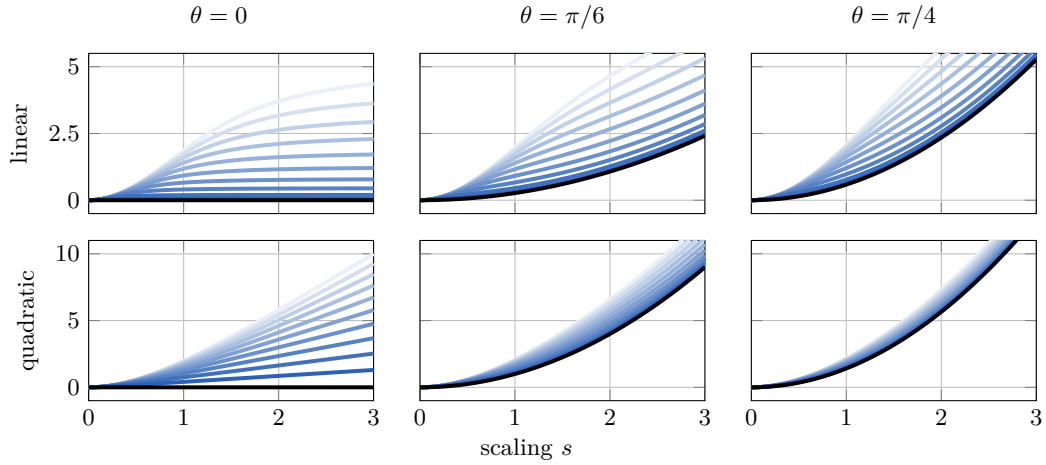


Figure 5.6. *Effect of scaling* both *vectors on the parallel level sets functionals. The regularization functions for linear parallel level sets* (5.39) *(top row) and quadratic parallel level sets* (5.40) *(bottom row) are shown as a function of the scaling s applied to both images. These are plotted for $t = 2s$ and $\theta = 0, \pi/6, \pi/4$. In each axis, the graphs for $\beta = 0$ are plotted in black and $\beta = 0.2, 0.4, \ldots, 2$ with decreasing intensity. See the text for more information.*

of the vector magnitudes s, t

$$f_{\mathrm{PL}_\beta^{\mathcal{L}}}(s,t) = \left[\beta^4 + \beta^2(s^2+t^2) + s^2t^2\right]^{1/2} - \left[\cos^2(\theta)s^2t^2 + \beta^4\right]^{1/2} \tag{5.39}$$

and

$$f_{\mathrm{PL}_\beta^{\mathcal{Q}}}(s,t) = \left[\beta^4 + \beta^2(s^2+t^2) + (1-\cos^2(\theta))s^2t^2\right]^{1/2} - \beta^2. \tag{5.40}$$

These two functions are plotted in figure 5.5 for fixed $t = 1$. For $\beta = 0$, shown

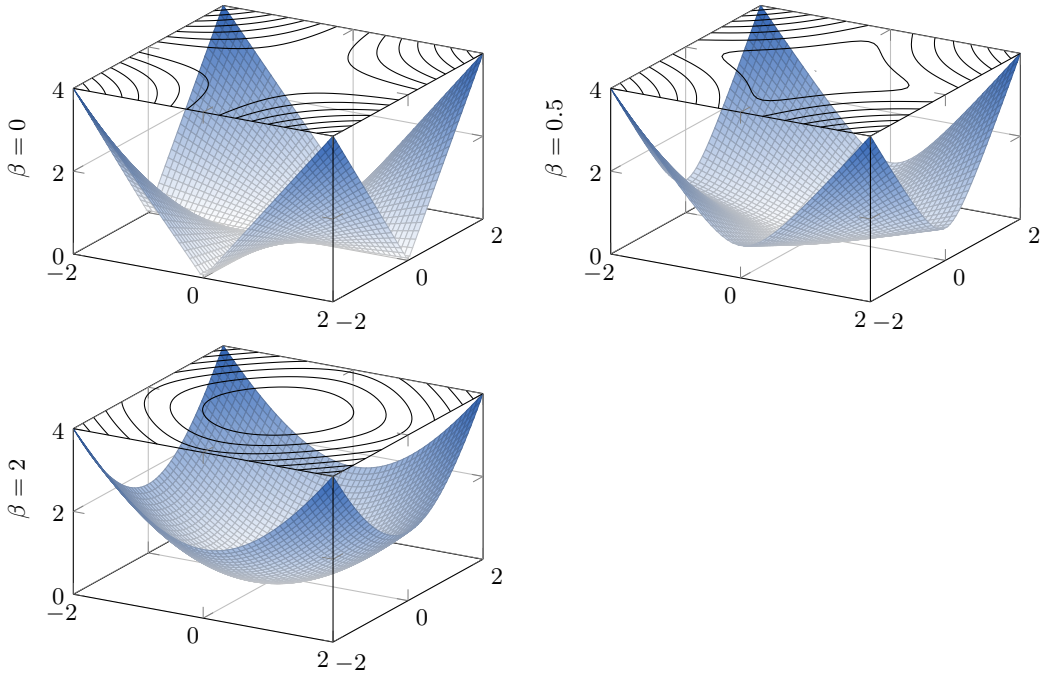


Figure 5.7. *The local regularization function for parallel level sets as a function of two vectors* (5.2.2)*. The plots are for orthogonal vectors so that linear and quadratic parallel level sets coincide. The parameter β changes the shape of the function locally around zero and makes the level sets more circular.*

in black, these are constant for $\theta = 0$ and linearly increasing for $\theta > 0$. In the case of $\beta > 0$, they all have linear growth for large s. While quadratic parallel level sets has a quadratic decay for small s, linear parallel level sets has this decay only for large β. For small but non-zero β, linear parallel level sets has a smaller function value for small but non-zero values than at zero.

The scaling of both functions simultaneously, *i.e.* $t = 2s$, is shown in figure 5.6. For $\theta > 0$ both functions have a quadratic growth for large s. When $\theta = 0$ and $\beta = 0$ both functions are constant zero. If $\beta > 0$ their behaviour is quite different. While linear parallel level sets grows first quadratically but then levels off for large s, quadratic parallel level sets has a linear growth.

Finally, we investigate the behaviour for scaling s and t independently. The case $\theta = \pi/2$ – where both functions are the same – is shown in figure 5.7. Although the functions are the same in each quadrant we plot the functions also for negative s and t for illustration purposes. When $\beta = 0$, the function is zero on the axis and non-zero everywhere else. For larger β, the level sets are rounded off and the functions become bowl shaped locally around zero. ▲

5.4. Comparison between Asymmetric Parallel Level Sets and Kaipio.

Proposition 5.4.1. *The model of Kaipio et al., i.e.* (5.33) *with matrix* (5.34)*, is a special case of asymmetric parallel level sets* (5.29) *with the choice* $\varphi = id, \psi(s) = \frac{1}{2}s^2$ *and* $\beta = 0$.

Proof. This is a direct consequence of lemma 5.4.3 later in this section. ∎

Example 5.4.2. The proposition says that the model of Kaipio et al. [86] is a special case of asymmetric parallel level sets. In particular, as it will become clear in the proof of lemma 5.4.3, the functions that normalize the side information η and $\tilde{\eta}$ are related in a one-to-one fashion. In this example, we shall compute the functions η and $\tilde{\eta}$ for asymmetric parallel level sets and for Kaipio's model explicitly.

In their paper, the function they use is

$$\tilde{\eta}(s) = (s^2 + \gamma^2)^{1/2} \tag{5.41}$$

with $\gamma = 1$ but we shall be a bit more general here and consider any $\gamma > 0$. According to the proof of lemma 5.4.3, this relates to

$$\eta(s) = (s^2 + \gamma^2)(s^2 + 2\gamma^2)^{-1/2}. \tag{5.42}$$

in the asymmetric parallel level sets model.

Conversely, the choice $\eta(s) = (s^2 + \gamma^2)^{1/2}$ in the asymmetric parallel level sets model corresponds to

$$\tilde{\eta}(s) = \left[s^2 + \gamma^2 + \gamma(s^2 + \gamma^2)^{1/2}\right]^{1/2} \tag{5.43}$$

in the model of Kaipio et al. For $\gamma = 0$ all these formulas coincide with the absolute value function but for $\gamma > 0$ they are different, see figure 5.8. ▲

Lemma 5.4.3. *Let* $\mathsf{X} \coloneqq \{\eta \colon [0,\infty) \to [0,\infty) \mid \eta \geq id, \eta > 0\}$. *For all* $\eta \in \mathsf{X}$ *there exists an* $\tilde{\eta} \in \mathsf{X}$ *such that for all* $x, y \in \mathbb{R}^D$*, we have*

$$|x|^2 - \langle x, y_\eta \rangle^2 = |Ax|^2 \tag{5.44}$$

with $A = I - y_{\tilde{\eta}} y_{\tilde{\eta}}^T$ *and vice versa.*

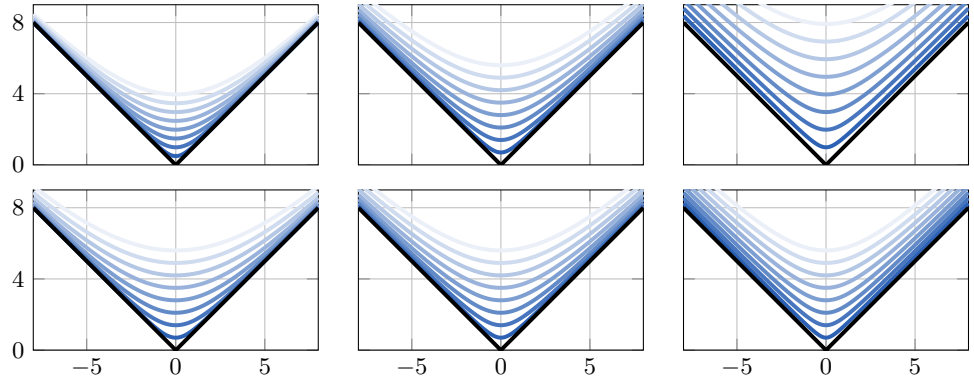


Figure 5.8. *Comparison of smoothed norms of asymmetric parallel level sets and Kaipio from example 5.4.2. From left to right:* (5.42), (5.41) *and* (5.43). *In the top row these are plotted for* $\gamma = 0.7, 1.4, \ldots, 5.6$ *with decreasing colour intensity. In the bottom row* γ *is scaled such that all three functions have the same value at 0. In addition, the absolute value function is shown in black.*

Proof. First, define $f, g \colon \mathsf{X} \to \mathsf{X}$ with

$$f_{\tilde{\eta}} \coloneqq f(\tilde{\eta}) \coloneqq \tilde{\eta}\left[2 - (id/\tilde{\eta})^2\right]^{-1/2} = \tilde{\eta}^2\left(2\tilde{\eta}^2 - id^2\right)^{-1/2} \tag{5.45}$$

$$g_{\eta} \coloneqq g(\eta) \coloneqq \eta\Big\{1 + \left[1 - (id/\eta)^2\right]^{1/2}\Big\}^{1/2} = \left[\eta^2 + \eta(\eta^2 - id^2)^{1/2}\right]^{1/2}. \tag{5.46}$$

Let us show first, that these two functions are well defined. Indeed, it is easy to see that $f_{\tilde{\eta}}(0) = \tilde{\eta}(0)/\sqrt{2} > 0$ and $g_{\eta}(0) = \sqrt{2}\eta(0) > 0$.

Next, denote $c \coloneqq \left[2 - (id/\tilde{\eta})^2\right]^{1/2} \geq 1$ and consider

$$f_{\tilde{\eta}} - id = \frac{\tilde{\eta}}{c} - id = \frac{\tilde{\eta} - \left[2 - (id/\tilde{\eta})^2\right]^{1/2} id}{c} \tag{5.47}$$

$$= \frac{\tilde{\eta}^2 - \left[2 - (id/\tilde{\eta})^2\right] id^2}{c(\tilde{\eta} + c)} = \frac{\left(\tilde{\eta}^2 - id^2\right)^2}{\tilde{\eta}^2 c(\tilde{\eta} + c)} \geq 0 \tag{5.48}$$

where we used that $\tilde{\eta} > 0$. As

$$g_{\eta} = \eta\Big\{1 + \left[1 - (id/\eta)^2\right]^{1/2}\Big\}^{1/2} \geq \eta \geq id, \tag{5.49}$$

we have shown that both functions are well-defined.

For the main part of the proof, let $\tilde{\eta} \in \mathsf{X}$. Then, with $f(\tilde{\eta})^{-2} = \tilde{\eta}^{-2}[2 - (id/\tilde{\eta})^2]$ and omitting the argument $|y|$, we have

$$|Ax|^2 = \left|(I - \tilde{\eta}^{-2} y y^T)x\right|^2 = \left|x - \tilde{\eta}^{-2} y \left\langle y, x\right\rangle\right|^2 \tag{5.50}$$

$$= |x|^2 - 2\tilde{\eta}^{-2}\left\langle y, x\right\rangle^2 + \tilde{\eta}^{-4}\left\langle y, x\right\rangle^2 |y|^2 \tag{5.51}$$

$$= |x|^2 - \tilde{\eta}^{-2}[2 - (|y|/\tilde{\eta})^2]\left\langle y, x\right\rangle^2 = |x|^2 - \left\langle x, y_{f(\tilde{\eta})}\right\rangle^2. \tag{5.52}$$

Finally, we see that g inverts f which completes the proof. Indeed, let $c \coloneqq 2\tilde{\eta}^2 - id^2$ we have $f_{\tilde{\eta}} = \tilde{\eta}^2/c$ and so

$$g(f_{\tilde{\eta}})^2 = \frac{\tilde{\eta}^4}{c} + \frac{\tilde{\eta}^2}{\sqrt{c}}\left(\frac{\tilde{\eta}^4}{c} - id^2\right)^{1/2} = \frac{\tilde{\eta}^4}{c} + \frac{\tilde{\eta}^2}{\sqrt{c}}\left(\frac{\tilde{\eta}^4 - 2\tilde{\eta}^2 id^2 + id^4}{c}\right)^{1/2} \tag{5.53}$$

$$= \frac{\tilde{\eta}^4}{c} + \frac{\tilde{\eta}^2(\tilde{\eta}^2 - id^2)}{c} = \frac{2\tilde{\eta}^4 - \tilde{\eta}^2 id^2}{c} = \tilde{\eta}^2. \tag{5.54}$$

■

Remark 5.4.4. As figure 5.8 indicates it seems possible to show that the function f in the proof is also a bijection on E, cf. (5.18). ▲

5.5. Jacobian Regularization and its Relation to Parallel Level Sets.

In this section, we will investigate another possibility to phrase a prior on vector-valued images. Instead of treating the gradients of each channel separately, we can view such a prior as a function of its *Jacobian*. As such it naturally leads to priors using the singular values of the Jacobian such as the Schatten p-norm. We will see that many of the aforementioned priors like separate total variation, joint total variation, quadratic parallel level sets and the cross-gradients functional are all special cases of this formalism.

5.5.1. Formulation with Jacobian. Another possibility to define a prior on vector-valued images is by

$$\mathcal{P}(w) = \int_\Omega f(Dw(x))\,\mathrm{d}x = \int_\Omega f(Dw). \tag{5.55}$$

Dw is the *Jacobian* of the vector-valued image $w\colon \Omega \subset \mathbb{R}^D \to \mathbb{R}^M$, i.e.

$$Dw(x) \coloneqq [\nabla w_1(x); \ldots; \nabla w_M(x)] \in \mathbb{R}^{D\times M} \tag{5.56}$$

where we treat the gradient as a column vector and adopt the MATLAB convention of concatenating vectors to form a matrix.

Remark 5.5.1. This notation on the Jacobian is equivalent to the previous notation on the gradient of the channels of a vector-valued image. For example, separate total variation and joint total variation fit into this framework by choosing

$$f(A) = \sum_{m=1}^{M}\left(\sum_{d=1}^{D} A_{d,m}^2\right)^{1/2} \quad \text{and} \quad f(A) = \left(\sum_{m=1}^{M}\sum_{d=1}^{D} A_{d,m}^2\right)^{1/2}, \tag{5.57}$$

respectively. ▲

5.5.2. Formulation with Singular Values. Beside this, many methods like separate total variation, joint total variation, quadratic parallel level sets and cross-gradients can be viewed as a special case of

$$\mathcal{P}(w) = \int_\Omega f(\sigma_1(x), \dots, \sigma_M(x)) \, \mathrm{d}x = \int_\Omega f(\sigma_1, \dots, \sigma_M) \tag{5.58}$$

where σ_m is the mth singular value of the Jacobian Dw. The function f is often chosen to be a *Schatten p-norm*, i.e. a p-norm on the vector of singular values. Examples that have been used in image processing include the 1-norm (total nuclear variation) [80, 133], 2-norm (joint total variation) [135, 74] and maximum-norm ($p = \infty$) [67].

5.5.3. Explicit Computation of the Singular Values. The singular values of Dw are the square-roots of the eigenvalues of the positive-semidefinite matrix

$$Dw^T Dw = \begin{pmatrix} |\nabla w_1|^2 & \langle \nabla w_1, \nabla w_2 \rangle & \dots & \langle \nabla w_1, \nabla w_M \rangle \\ \langle \nabla w_1, \nabla w_2 \rangle & |\nabla w_2|^2 & \ddots & \langle \nabla w_2, \nabla w_M \rangle \\ \vdots & \ddots & \ddots & \vdots \\ \langle \nabla w_1, \nabla w_M \rangle & \langle \nabla w_2, \nabla w_M \rangle & \dots & |\nabla w_M|^2 \end{pmatrix}. \tag{5.59}$$

In the case $M = 1$, i.e. there is only one modality, denote $u = w_1$. The only singular value is $|\nabla u|$ so every Schatten p-norm on the Jacobian is a generalization of total variation to vector-valued images. In fact, every function $f\colon [0, \infty)^M \to \mathbb{R}$ that reduces to the identity for $M = 1$ has this property.

For $M = 2$, i.e. we consider two modalities, let $(u, v) = w$. We need to compute the square root of the roots of the determinant of

$$\begin{pmatrix} |\nabla u|^2 - \sigma^2 & \langle \nabla u, \nabla v \rangle \\ \langle \nabla u, \nabla v \rangle & |\nabla v|^2 - \sigma^2 \end{pmatrix}. \tag{5.60}$$

These are

$$\sigma_{1,2} = \left[m \pm (m^2 - d)^{1/2} \right]^{1/2} \tag{5.61}$$

with the mean of the squared norms $m = 1/2(|\nabla u|^2 + |\nabla v|^2)$ and the determinant $d = |\nabla u|^2 \, |\nabla v|^2 - \langle \nabla u, \nabla v \rangle^2$.

Table 5.1. *Different methods and their expression in terms of singular values of the Jacobian. In order to generalize TV the method has to reduce to a function of σ_1 if $\sigma_2 = 0$.*

	expression with singular values	aligns gradients	generalizes TV
Schatten 1-norm [80, 133]	$\sigma_1 + \sigma_2$	yes	yes
Schatten 2-norm, $\mathbb{TV}^{\mathcal{J}}$ [135, 74]	$\left(\sigma_1^2 + \sigma_2^2\right)^{1/2}$	no	yes
Schatten ∞-norm [67]	σ_1	no	yes
cross-gradients [63, 74]	$\sigma_1^2\sigma_2^2$	yes	no
$\mathrm{PL}^{\mathcal{Q}}$	$\sigma_1\sigma_2$	yes	no
$\mathrm{PL}_{\beta}^{\mathcal{Q}}$ [139, 51]	$\left(\beta^4 + \beta^2(\sigma_1^2 + \sigma_2^2) + \sigma_1^2\sigma_2^2\right)^{1/2}$	yes	yes

Remark 5.5.2. Note that this derivation depends on the number of modalities M but not on the dimensions of the underlying images D. ▲

Remark 5.5.3. The cross-gradient functional, see example 5.2.6, and used in [63, 64, 65, 7, 110, 74] coincides with the determinant d. ▲

Remark 5.5.4. If the gradients align, i.e. $\nabla u \parallel \nabla v$, then the determinant vanishes and the singular values reduce to

$$\sigma_1 = \left(|\nabla u|^2 + |\nabla v|^2\right)^{1/2} \quad \text{and} \quad \sigma_2 = 0. \tag{5.62}$$

▲

We have seen that several methods can be written in terms of the singular values of the Jacobian. An overview of the methods in terms of singular values is given in table 5.1. As both joint total variation and cross-gradients are among them also quadratic parallel level sets can be written in terms of singular values.

5.6. Summary.

We have introduced the general concept of parallel level sets that aims to align the spatial gradients of images from different modalities. In particular, we introduced the symmetric and asymmetric parallel level sets functionals that both depend on the very same underlying notion. Within these classes, different choices of φ and ψ lead to members with very different properties. Examples include linear and quadratic parallel level sets.

Chapter 6

Gâteaux Derivative and Diffusive Flow of Parallel Level Sets

We have seen in chapter 4 that the Gâteaux derivative of gradient based regularization functionals is connected to a diffusive flow. In this chapter, we will compute the Gâteaux derivative of the symmetric and asymmetric parallel level sets functional and show that this – in contrast to the methods in chapter 4 which result in isotropic diffusion – relates to *anisotropic diffusion*. The last part of this chapter is dedicated to the discretization of the Gâteaux derivative.

6.1. Gâteaux Derivative of Parallel Level Sets.

Before we can analyse the diffusion of parallel level sets, we need to compute the corresponding Gâteaux derivative first.

6.1.1. (Symmetric) Parallel Level Sets.

Theorem 6.1.1. *Let $\beta > 0$ and $\tilde{\psi} := \psi'/id$. In addition, let the matrix-valued mapping $\varkappa(x,y) = a(x,y)I - b(x,y)yy^T$ be given with scalar coefficients*

$$a(x,y) = |y|_\beta^2 \, \tilde{\psi}(|x|_\beta \, |y|_\beta)\varphi'[\psi(|x|_\beta \, |y|_\beta) - \psi(|\langle x,y\rangle|_{\beta^2})] \tag{6.1}$$

$$b(x,y) = \tilde{\psi}(|\langle x,y\rangle|_{\beta^2})\varphi'[\psi(|x|_\beta \, |y|_\beta) - \psi(|\langle x,y\rangle|_{\beta^2})]. \tag{6.2}$$

Then the Gâteaux derivatives for symmetric parallel level sets, cf. (5.17)*, can be written in the generic form of diffusion*

$$D\,\mathrm{PL}_\beta[u] = -\operatorname{div}(\varkappa(\nabla u, \nabla v)\nabla u) \tag{6.3}$$

$$D\,\mathrm{PL}_\beta[v] = -\operatorname{div}(\varkappa(\nabla v, \nabla u)\nabla v). \tag{6.4}$$

We will prove this statement at the end of this section but let us first have a look at the special cases considered before. These derivatives are not really instructive in itself and primarily given as a reference for the diffusive flow later in this chapter.

Example 6.1.2 (Linear Parallel Level Sets). For linear parallel level sets, i.e. $\varphi, \psi = id$, the scalar coefficients are given by

$$a(x,y) = |y|_\beta \, |x|_\beta^{-1} \quad \text{and} \quad b(x,y) = |\langle x,y\rangle|_{\beta^2}^{-1} . \tag{6.5}$$

▲

Example 6.1.3 (Quadratic Parallel Level Sets). The scalar coefficients for quadratic parallel level sets are

$$a(x,y) = |y|_\beta^2 \left[\beta^4 + \beta^2(|x|^2 + |y|^2) + |x|^2 \, |y|^2 - |\langle x,y\rangle|^2 \right]^{-1/2} \tag{6.6}$$

$$b(x,y) = \left[\beta^4 + \beta^2(|x|^2 + |y|^2) + |x|^2 \, |y|^2 - |\langle x,y\rangle|^2 \right]^{-1/2} . \tag{6.7}$$

▲

Example 6.1.4 (Cross-Gradients). For cross-gradients, see example 5.2.6, the coefficients are given by

$$a(x,y) = |y|^2 \quad \text{and} \quad b(x,y) = 1. \tag{6.8}$$

▲

Proof of theorem 6.1.1. In order to use theorem 4.2.5 to compute the Gâteaux derivatives for the parallel level sets functional, we need to compute the gradient of its local regularization function. Recall that the local regularization function for symmetric parallel level sets has been defined in (5.17) as

$$f_{\mathrm{PL}_\beta}(x,y) \coloneqq \varphi\Big[\psi(|x|_\beta \, |y|_\beta) - \psi(|\langle x,y\rangle|_{\beta^2})\Big]. \tag{6.9}$$

As the function is symmetric, it is sufficient to compute its gradient with respect to x. We will make use of the multi-dimensional chain rule [47, p. 51] and the derivatives of both the smoothed norm and the Euclidean scalar product. The latter are given by $\nabla |x|_\beta = x/|x|_\beta$ and $\nabla_x \langle x,y\rangle = y$. In addition, we use the trivial but important equality

$$\underbrace{\langle x,y\rangle}_{\in\mathbb{R}} \underbrace{y}_{\in\mathbb{R}^D} = y\,\langle y,x\rangle = \underbrace{yy^T}_{\in\mathbb{R}^{D\times D}} \underbrace{x}_{\in\mathbb{R}^D}. \tag{6.10}$$

Let $c = \varphi'[\psi(|x|_\beta\,|y|_\beta) - \psi(|\!\langle x,y\rangle\!|_{\beta^2})]$. Then the gradient of f_{PL_β} is given by

$$\nabla_x f_{\mathrm{PL}_\beta}(x,y) = c\Big[\frac{\psi'(|x|_\beta\,|y|_\beta)\,|y|_\beta}{|x|_\beta}x - \frac{\psi'(|\!\langle x,y\rangle\!|_{\beta^2})}{|\!\langle x,y\rangle\!|_{\beta^2}}\,\langle x,y\rangle\,y\Big] \tag{6.11}$$

$$= c\Big[\tilde{\psi}(|x|_\beta\,|y|_\beta)\,|y|_\beta^2\,I - \tilde{\psi}(|\!\langle x,y\rangle\!|_{\beta^2})yy^T\Big]x \tag{6.12}$$

and the application of theorem 4.2.5 completes the proof. ■

Remark 6.1.5. The previous proof, cf. (6.11), shows that the Gâteaux derivative of the parallel level set functional can also be written as

$$D\,\mathrm{PL}_\beta[u] = -\operatorname{div}(\kappa\nabla u) - \operatorname{div}(\tau\nabla v) \tag{6.13}$$

with scalar κ and τ. This simplifies the numerical calculations and we will make use of this formulation in the numerical examples. ▲

6.1.2. Asymmetric Parallel Level Sets. We will now compute the Gâteaux derivatives for the asymmetric parallel level sets functional. As it is asymmetric, we have to compute both derivatives separately.

Theorem 6.1.6. *Let $\beta > 0$ and $\tilde{\psi} := \psi'/id$. Moreover, let*

$$\varkappa_1 = a_1 I - b_1 y_\eta y_\eta^T \qquad\qquad \varkappa_2 = a_2 I - b_2 x x^T \tag{6.14}$$

with

$$a_1 = c\tilde{\psi}(|x|_\beta) \qquad\qquad b_1 = c\tilde{\psi}(|\!\langle x,y_\eta\rangle\!|_\beta) \tag{6.15}$$

$$a_2 = c\tilde{\psi}(|\!\langle x,y_\eta\rangle\!|_\beta)\,\langle x,y_\eta\rangle^2\,\frac{\eta'(|y|)}{\eta(|y|)\,|y|} \qquad b_2 = c\tilde{\psi}(|\!\langle x,y_\eta\rangle\!|_\beta)\eta(|y|)^{-2} \tag{6.16}$$

$$c = \varphi'[\psi(|x|_\beta) - \psi(|\!\langle x,y_\eta\rangle\!|_\beta)]. \tag{6.17}$$

Then the Gâteaux derivatives for asymmetric parallel level sets are given by

$$D\,\mathrm{APL}_\beta[u] = -\operatorname{div}(\varkappa_1\nabla u) \tag{6.18}$$

$$D\,\mathrm{APL}_\beta[v] = -\operatorname{div}(\varkappa_2\nabla v). \tag{6.19}$$

Example 6.1.7 (Linear Asymmetric Parallel Level Sets). In the case of linear asymmetric parallel level sets, i.e. $\varphi, \psi = id, \beta > 0$, the coefficients of $\varkappa_n$ are

$$a_1 = |x|_\beta^{-1} \qquad b_1 = |\!\langle x, y_\eta \rangle\!|_\beta^{-1} \tag{6.20}$$

$$a_2 = |\!\langle x, y_\eta \rangle\!|_\beta \frac{\eta'(|y|)}{\eta(|y|)\,|y|} \qquad b_2 = |\!\langle x, y_\eta \rangle\!|_\beta^{-1}\, \eta(|y|)^{-2}. \tag{6.21}$$

▲

Example 6.1.8 (Quadratic Asymmetric Parallel Level Sets). The coefficients of the diffusivities for quadratic parallel level sets, i.e. $\varphi(s) = (\beta^2 + s)^{1/2}, \psi(s) = s^2, \beta > 0$, are

$$a_1 = c \qquad b_1 = c \tag{6.22}$$

$$a_2 = c\,\langle x, y_\eta \rangle^2 \frac{\eta'(|y|)}{\eta(|y|)\,|y|} \qquad b_2 = c\eta(|y|)^{-2} \tag{6.23}$$

$$c = \left(\beta^2 + |x|^2 - \langle x, y_\eta \rangle^2\right)^{-1/2}. \tag{6.24}$$

▲

Example 6.1.9 (Kaipio). For the model of Kaipio et al.[86], i.e. $\varphi = id, \psi(s) = \frac{1}{2}s^2, \beta = 0$, the scalar coefficients of the matrix-valued diffusivities are given by

$$a_1 = 1 \qquad b_1 = 1 \tag{6.25}$$

$$a_2 = \langle x, y_\eta \rangle^2 \frac{\eta'(|y|)}{\eta(|y|)\,|y|} \qquad b_2 = \eta(|y|)^{-2}. \tag{6.26}$$

▲

To complete this section, we prove theorem 6.1.6.

Proof of theorem 6.1.6. Recall that the local regularization function of asymmetric parallel level sets (5.29) is given by

$$f_{\text{APL}}(x; y) \coloneqq \varphi\Big[\psi(|x|_\beta) - \psi(|\!\langle x, y_\eta \rangle\!|_\beta)\Big]. \tag{6.27}$$

Let $c = \varphi'\left[\psi(|x|_\beta) - \psi(|\!\langle x, y_\eta \rangle\!|_\beta)\right]$. Then the gradient with respect to the first argument is

$$\nabla_x f_{\text{APL}}(x; y) = c\Big\{\frac{\psi'(|x|_\beta)}{|x|_\beta} x - \frac{\psi'(|\!\langle x, y_\eta \rangle\!|_\beta)}{|\!\langle x, y_\eta \rangle\!|_\beta} \langle x, y_\eta \rangle\, y_\eta\Big\} \tag{6.28}$$

$$= c\Big\{\tilde{\psi}(|x|_\beta) I - \tilde{\psi}(|\!\langle x, y_\eta \rangle\!|_\beta) y_\eta y_\eta^T\Big\} x \tag{6.29}$$

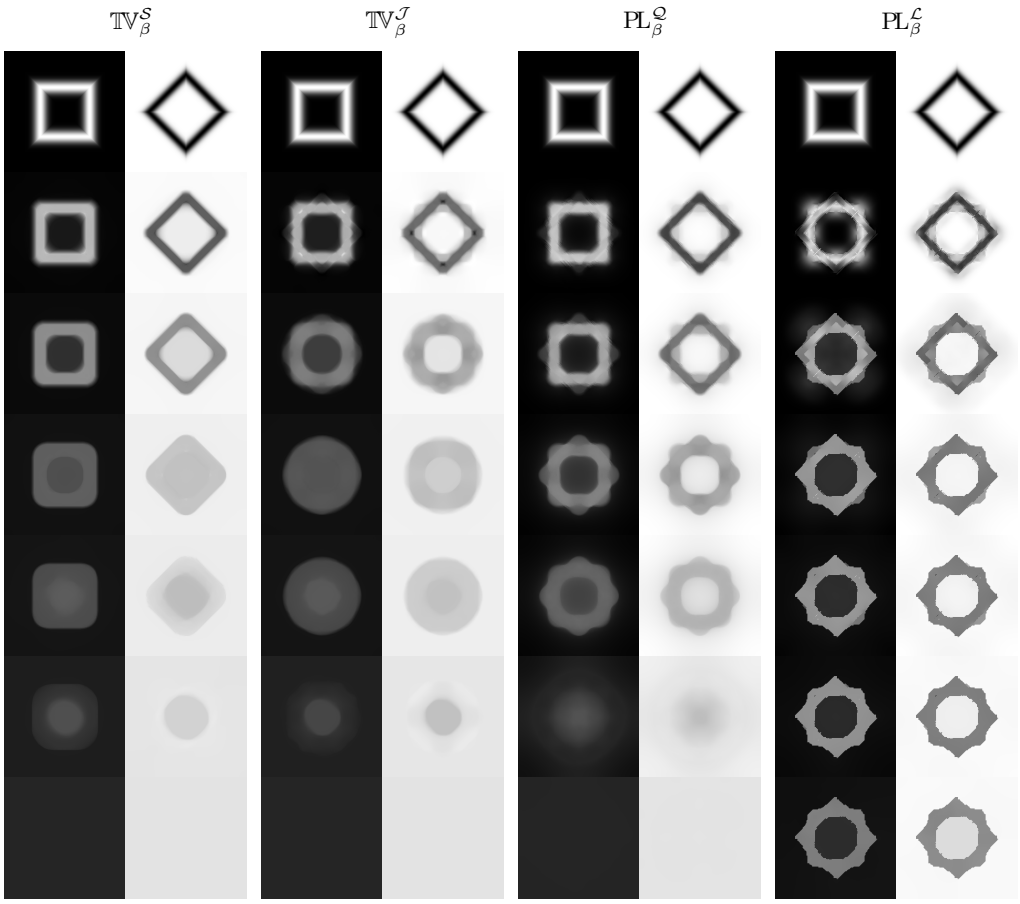


Figure 6.1. *Minimizing sequences for total variation, joint total variation, quadratic and linear parallel level sets. During the iterations from top to bottom images with lower function value are obtained. While $\mathbb{TV}^{\mathcal{S}}_{\beta}$ (β=1e-4), $\mathbb{TV}^{\mathcal{J}}_{\beta}$ (β=1e-4) and $\mathrm{PL}^{\mathcal{Q}}_{\beta}$ (β=3e-3) converge to a constant image $\mathrm{PL}^{\mathcal{L}}_{\beta}$ (β=5e-2) has a non-trivial stationary point. The images are obtained with a Quasi-Newton method (L-BFGS). Image values are in [0,1] and the grid size is chosen to be of unit length.*

and with respect to the second argument we have

$$\begin{aligned}\nabla_y f_{\mathrm{APL}}(x;y) &= c[-\psi'(|\langle x, y_\eta\rangle|_\beta)]\,|\langle x, y_\eta\rangle|_\beta^{-1}\,\langle x, y_\eta\rangle \\ &\qquad \Big\{\eta(|y|)^{-1}x - \langle x, y\rangle\,\eta(|y|)^{-2}\eta'(|y|)\,|y|^{-1}\,y\Big\} \qquad (6.30)\\ &= c\tilde{\psi}(|\langle x, y_\eta\rangle|_\beta)\Big\{\langle x, y_\eta\rangle^2\,\frac{\eta'(|y|)}{\eta(|y|)\,|y|}I - \eta(|y|)^{-2}xx^T\Big\}y. \qquad (6.31)\end{aligned}$$

■

6.2. Analysis of Diffusive Flow.

In the previous section we computed the Gâteaux derivative of symmetric and asymmetric parallel level sets. Using these derivatives we illustrate some properties of total variation, joint total variation, quadratic and linear parallel level sets. Figure 6.1 shows four minimizing sequences for these priors, i.e. from top to bottom the images have a decreasing function value. While the image pairs for total variation have independent

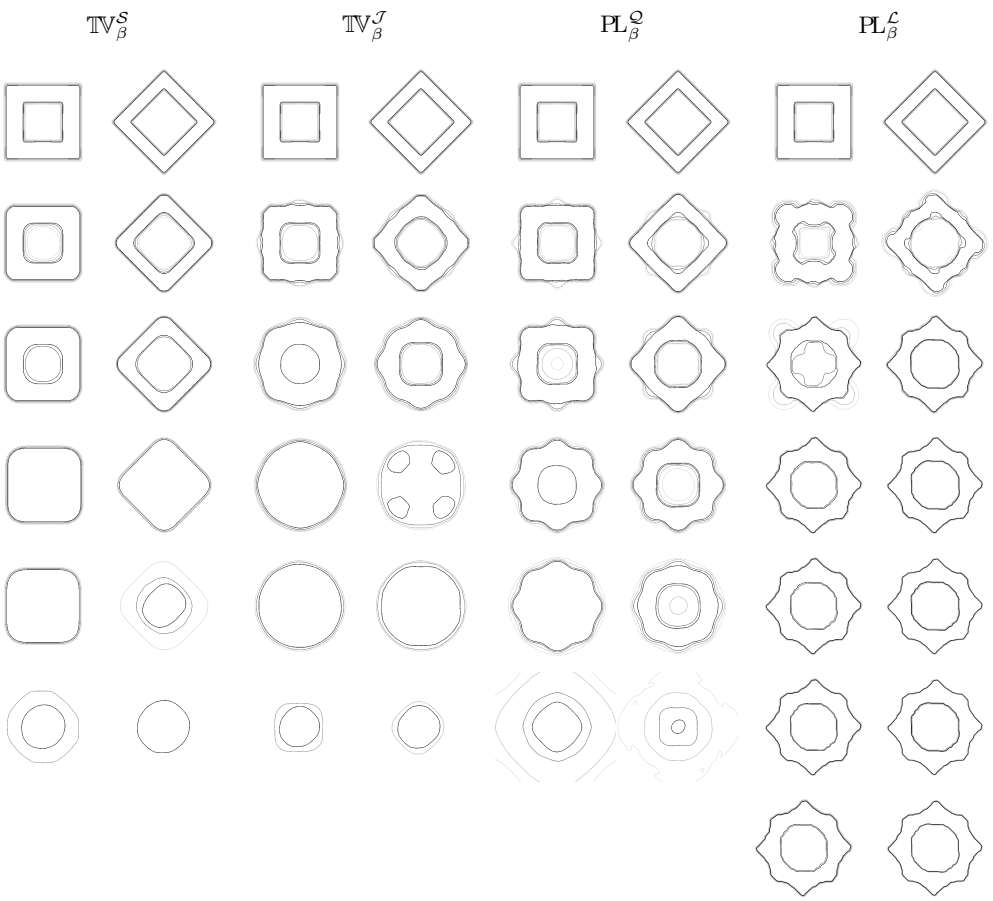


Figure 6.2. *Level sets of minimizing sequences for total variation, joint total variation, quadratic and linear parallel level sets shown in figure 6.1. From top to bottom are the level sets of the iterates for an increasing number of iterations.*

shapes, for the other three priors the shapes are correlated. This is highlighted in figure 6.2 which shows the corresponding level sets. It is also interesting to see that the total variation based priors and quadratic parallel level sets converge to constant images with the constant being their mean value. In contrast, the linear parallel level sets prior has a non-trivial stationary point where the images are piecewise constant with similar shapes. Similarly, figure 6.4 shows that the minimizing sequences of total variation, joint total variation and quadratic parallel level sets have function values converging to zero whereas the function values of linear parallel level sets stay larger than two. The global optimum is known to be zero for all four priors so only the former group have sequences converging to a global optimum.

In this section we will analyse the parallel level set priors with respect to their diffusivities and explain thereby some of the aforementioned observations.

For the analysis of the diffusivity we need to remind the reader about the orthogonal complement and the span.

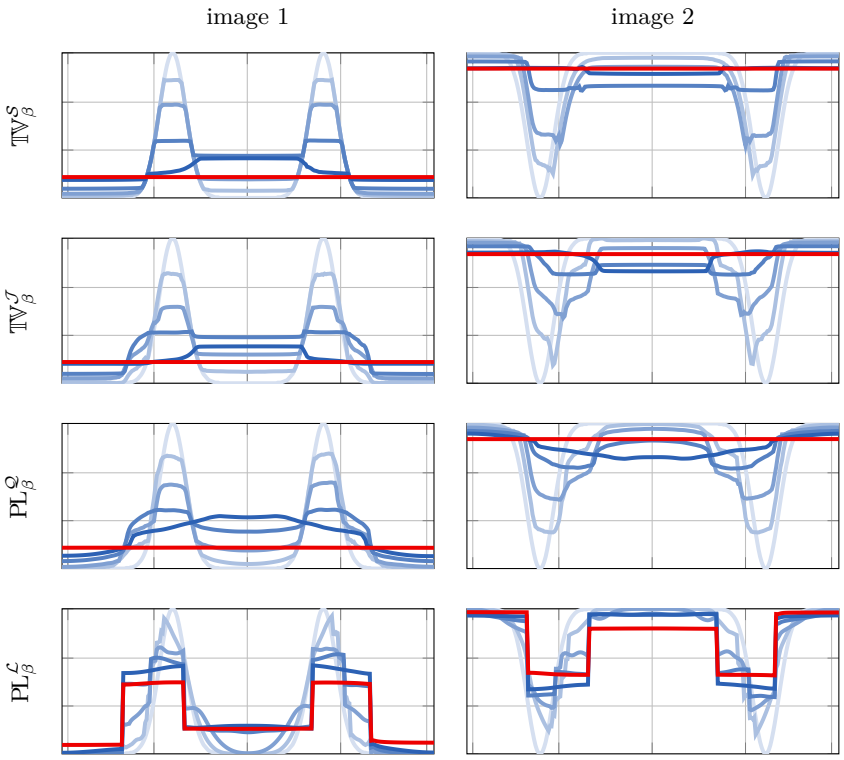


Figure 6.3. *Line profiles of minimizing sequences shown in figure 6.1. In each axes the line profiles for an increasing number of iterations are shown with increasing colour intensity. The stationary profile is shown in red.*

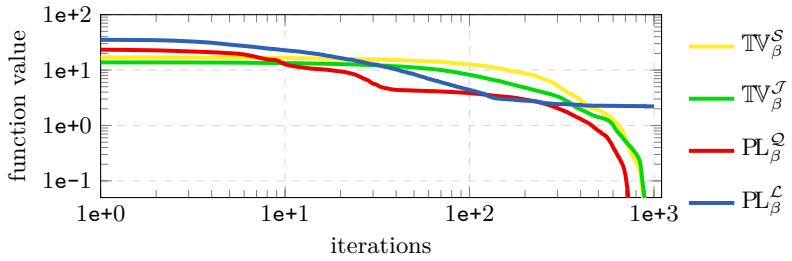


Figure 6.4. *Function values of minimizing sequences shown in figure 6.1. As the priors are scaled such that the global solution has a function value of zero this figure shows that while total variation, joint total variation and quadratic parallel level sets converge to a global minimizer, linear parallel level sets does not.*

Definition 6.2.1 (Span, Orthogonal Complement). *For any $y \in \mathbb{R}^D$ we can decompose* $\mathbb{R}^D = \operatorname{span}(y) \oplus \operatorname{orth}(y)$ *where the* span *and the* orthogonal complement *of y are defined as*

$$\operatorname{span}(y) := \left\{ sy \in \mathbb{R}^D \mid s \in \mathbb{R} \right\} \quad \textit{and} \quad \operatorname{orth}(y) := \left\{ x \in \mathbb{R}^D \mid \langle x, y \rangle = 0 \right\}. \tag{6.32}$$

To analyse the diffusion we need to introduce *local coordinates* of an image domain in terms of the level set normals and tangents. Remember that the image gradient is perpendicular to the level sets. These coordinates are also called *Gauge coordinates*

[158, 4].

Definition 6.2.2 (Local Coordinates). *Let* $u\colon \Omega \subset \mathbb{R}^D \to \mathbb{R}$ *be a differentiable function. We denote by* R_u *the* local coordinates $R_u\colon \Omega \to \mathbb{R}^{D\times D}$ *such that for any point* $x \in \Omega$ *the columns of* $R_u(x)$ *form a basis of* $\operatorname{span}(\nabla u(x)) \oplus \operatorname{orth}(\nabla u(x))$.

6.2.1. Diffusion Analysis for Parallel Level Sets. There is everything in place to state one of the main results of this thesis.

Theorem 6.2.3. *Let* $\beta > 0, \tilde{\psi} \coloneqq \psi'/id$ *and* R_u, R_v *be local coordinates for* u *and* v. *Moreover, let* $\Lambda(x,y) \coloneqq \operatorname{diag}(\lambda^{\perp}(x,y), \lambda^{\parallel}(x,y), \dots, \lambda^{\parallel}(x,y))$ *with*

$$\lambda^{\perp}(x,y) \coloneqq c\Big[\tilde{\psi}(|x|_\beta\,|y|_\beta)\,|y|_\beta^2 - \tilde{\psi}(|\langle x,y\rangle|_{\beta^2})\,|y|^2\Big] \tag{6.33}$$

$$\lambda^{\parallel}(x,y) \coloneqq c\tilde{\psi}(|x|_\beta\,|y|_\beta)\,|y|_\beta^2 \tag{6.34}$$

$$c \coloneqq \varphi'\big[\psi(|x|_\beta\,|y|_\beta) - \psi(|\langle x,y\rangle|_{\beta^2})\big]. \tag{6.35}$$

Then the Gâteaux derivatives computed in theorem 6.1.1 can be written in terms of their principal directions as

$$D\,\mathrm{PL}_\beta[u] = -\operatorname{div}\big(R_v\,\Lambda(\nabla v, \nabla u)\,R_v^T\,\nabla u\big) \tag{6.36}$$

$$D\,\mathrm{PL}_\beta[v] = -\operatorname{div}\big(R_u\,\Lambda(\nabla v, \nabla u)\,R_u^T\,\nabla v\big). \tag{6.37}$$

Remark 6.2.4. Theorem 6.2.3 says that the diffusive flow for an image u generated by the parallel level sets functional has principal directions given by the other image v and vice versa. Moreover, the directions but not the amount of diffusion are independent of the functions φ and ψ. ▲

Remark 6.2.5. Both c and $\lambda^{\parallel}$ are non-negative. Indeed, the functions φ and ψ were chosen to be monotonically increasing, hence their derivative is non-negative and the claim follows from

$$|x|_\beta\,|y|_\beta = \Big[\beta^4 + \beta^2(|x|^2 + |y|^2) + |x|^2\,|y|^2\Big]^{1/2} \geq \big||x|\,|y|\big|_{\beta^2} \geq |\langle x,y\rangle|_{\beta^2}\,. \tag{6.38}$$

For $\beta = 0$ the sign of $\lambda^{\perp}$ depends on the monotonicity of $\tilde{\psi}$. ▲

Remark 6.2.6. In the case of $\beta = 0$, more assumptions on φ, ψ are needed rather than $\mathsf{C}^1([0,\infty))$ in order to ensure differentiability. ▲

The proof of the theorem is based on the following lemma on matrices.

Lemma 6.2.7. *Let $y \in \mathbb{R}^D, a, b \in \mathbb{R}$ and $\varkappa = aI - b\, yy^T \in \mathbb{R}^{D\times D}$. Moreover, let $\lambda^\perp := a - b\,|y|^2$ and $\lambda^\parallel := a$. Then the following holds true.*

1. *The matrix $\varkappa$ is symmetric.*
2. *All $x \in \operatorname{span}(y)$ are eigenvectors to the eigenvalue $\lambda^\perp$.*
3. *All $x \in \operatorname{orth}(y)$ are eigenvectors to the eigenvalue $\lambda^\parallel$.*

Proof. Ad 1) There is

$$\varkappa^T = (aI - b\, yy^T)^T = aI^T - b\,(yy^T)^T = aI - b\, yy^T = \varkappa. \tag{6.39}$$

Ad 2) Let $x \in \operatorname{span}(y)$. So there is $x = sy$ for some $s \in \mathbb{R}$ and hence

$$\varkappa x = s(aI - b\, yy^T)y = s(ay - b\,|y|^2\, y) = \lambda^\perp x. \tag{6.40}$$

Ad 3) Let $x \in \operatorname{orth}(y)$. Then we have

$$\varkappa x = (aI - b\, yy^T)x = ax - b\, y\,\langle y, x\rangle = ax = \lambda^\parallel x. \tag{6.41}$$

■

Proof of theorem 6.2.3. This result follows immediately from theorem 4.2.5 combined with lemma 6.2.7. ■

Let us have a look at the examples of the general parallel level set functional and analyse them by means of their principal diffusivities. Before we start the analysis we formulate some desirable properties.

Desirable Properties 6.2.8. *Desirable properties for a gradient based regularization that preserves both the structure in both images are:*

i) It reduces to isotropic diffusion when the side information vanishes, i.e. for $|y| \to 0$ there is $\lambda^\perp = \lambda^\parallel$

ii) It reduces to an edge-preserving regularization when the side information vanishes, e.g. for $|y| \to 0$ there is $\lambda^\perp, \lambda^\parallel \approx 1/|x|$.

iii) In case both gradients vanish, the diffusion becomes linear, homogeneous and isotropic, e.g. for $|x|, |y| \to 0$ there is $\lambda^\perp, \lambda^\parallel = 1$.

Example 6.2.9 (Linear Parallel Level Sets). The principal diffusivities of linear parallel level sets are given by

$$\lambda^{\perp}(x,y) = |y|_\beta \, |x|_\beta^{-1} - |\langle x,y\rangle|_{\beta^2}^{-1} \, |y|^2 \quad \text{and} \quad \lambda^{\parallel}(x,y) = |y|_\beta \, |x|_\beta^{-1} . \tag{6.42}$$

Let us check whether the linear parallel level set prior has the desirable properties listed in 6.2.8. For $y \to 0$ there is $\lambda^{\perp} = \lambda^{\parallel} = \beta \, |x|_\beta^{-1}$ which shows that properties i) and ii) are fulfilled. In addition, it is easy to see that if $x \to 0$ then $\lambda^{\perp} = \lambda^{\parallel} = 1$ so that also property iii) holds.

For $\beta = 0, |x| \, , |y| > 0$ and $x \parallel y$ there is $\lambda^{\perp} = |y| \, |x|^{-1} - |x|^{-1} \, |y|^{-1} \, |y|^2 = 0$. Let $y = sx$, then $\lambda^{\parallel} = |s|$ which converges to zero if $y \to 0$ or $|x| \to \infty$.

Let either $\beta = 0$ or $|y| \gg 0$ such that $|y| \, / \, |y|_\beta \approx 1$. Then there is

$$\lambda^{\perp} = |y|_\beta^2 \left[|y|_\beta^{-1} \, |x|_\beta^{-1} - |\langle x,y\rangle|_{\beta^2}^{-1} \, (|y| \, / \, |y|_\beta)^2 \right] \tag{6.43}$$

$$\approx |y|_\beta^2 \, (|y|_\beta^{-1} \, |x|_\beta^{-1} - |\langle x,y\rangle|_{\beta^2}^{-1}) \leq 0. \tag{6.44}$$

This shows that the diffusion across the edge can be negative. ▲

Example 6.2.10 (Quadratic Parallel Level Sets). The principal diffusivities of quadratic parallel level sets are

$$\lambda^{\perp}(x,y) = \beta^2 [\beta^4 + \beta^2(|x|^2 + |y|^2) + |x|^2 \, |y|^2 - \langle x,y\rangle^2]^{-1/2} \tag{6.45}$$

$$\lambda^{\parallel}(x,y) = |y|_\beta^2 \, [\beta^4 + \beta^2(|x|^2 + |y|^2) + |x|^2 \, |y|^2 - \langle x,y\rangle^2]^{-1/2}. \tag{6.46}$$

It is easy to see that in the case of vanishing side information, i.e. $y \to 0$, there is $\lambda^{\perp} = \lambda^{\parallel} = \beta \, |x|_\beta^{-1}$ and with the same arguments as for the linear case we see that it fulfils all three desirable properties.

In addition, let $\beta = 0$ and $x \to sy$ then $\lambda^{\perp}, \lambda^{\parallel} \to +\infty$. For non-zero β, quadratic parallel level sets has diffusivities similar to joint total variation for $x \to sy$. ▲

Remark 6.2.11. Example 6.2.9 shows that linear parallel level sets yields in some cases negative principal diffusivities. This is commonly referred to as *backward diffusion*. Backward diffusion can lead to instabilities but also might result in sharper images. The principal diffusivities of quadratic parallel level sets are all non-negative which makes it more stable. This analysis explains the observation in [50] where linear parallel sets resulted in sharper images compared to quadratic parallel level sets. ▲

Example 6.2.12 (Cross-Gradients). The principal diffusivities of the cross-gradients functional are given by

$$\lambda^{\perp}(x,y) = 0 \quad \text{and} \quad \lambda^{\parallel}(x,y) = |y|^2 . \tag{6.47}$$

The flow generated is very simple – there is diffusion for one image only along the edges of the other image and the diffusion stops when the gradient of the other image vanishes. ▲

6.2.2. Diffusion Analysis for Asymmetric Parallel Level Sets.

Theorem 6.2.13. *Let $\beta > 0$ and R_u, R_v be local coordinates for u and v. In addition, let $\Lambda_n \coloneqq \operatorname{diag}(\lambda_n^{\perp}, \lambda_n^{\parallel}, \dots, \lambda_n^{\parallel}), n = 1, 2$ with*

$$\lambda_1^{\perp} \coloneqq c_1 \Big[\tilde{\psi}(|x|_\beta) - \tilde{\psi}(|\langle x, y_\eta \rangle|_\beta)\, |y_\eta|^2 \Big] \qquad \lambda_1^{\parallel} \coloneqq c_1 \tilde{\psi}(|x|_\beta) \tag{6.48}$$

$$c_1 \coloneqq \varphi' [\psi(|x|_\beta) - \psi(|\langle x, y_\eta \rangle|_\beta)] \tag{6.49}$$

$$\lambda_2^{\perp} \coloneqq c_2 \Big[\langle x, y_\eta \rangle^2 \, \frac{\eta'(|y|)}{\eta(|y|)\,|y|} - \eta(|y|)^{-2} \, |x|^2 \Big] \qquad \lambda_2^{\parallel} \coloneqq c_2 \, \langle x, y_\eta \rangle^2 \, \frac{\eta'(|y|)}{\eta(|y|)\,|y|} \tag{6.50}$$

$$c_2 \coloneqq \varphi' [\psi(|x|_\beta) - \psi(|\langle x, y_\eta \rangle|_\beta)] \tilde{\psi}(|\langle x, y_\eta \rangle|_\beta). \tag{6.51}$$

Then the Gâteaux derivatives (6.18) *and* (6.19) *can be written as*

$$D\,\mathrm{APL}_\beta[u] = -\operatorname{div}(R_v \, \Lambda_1 \, R_v^T \, \nabla u) \tag{6.52}$$

$$D\,\mathrm{APL}_\beta[v] = -\operatorname{div}(R_u \, \Lambda_2 \, R_u^T \, \nabla v). \tag{6.53}$$

Proof. The claim follows again from a combination of theorem 6.1.6 with lemma 6.2.7. ■

Example 6.2.14 (Linear Asymmetric Parallel Level Sets). The principal diffusivities of linear parallel level sets are given by

$$\lambda_1^{\perp} = |x|_\beta^{-1} - |\langle x, y_\eta \rangle|_\beta^{-1} \, |y_\eta|^2 \qquad \lambda_1^{\parallel} = |x|_\beta^{-1} \tag{6.54}$$

$$\lambda_2^{\perp} = \frac{\langle x, y_\eta \rangle^2}{|\langle x, y_\eta \rangle|_\beta} \frac{\eta'(|y|)}{\eta(|y|)\,|y|} - |\langle x, y_\eta \rangle|_\beta^{-1} \, \eta(|y|)^{-2} \, |x|^2 \qquad \lambda_2^{\parallel} = \frac{\langle x, y_\eta \rangle^2}{|\langle x, y_\eta \rangle|_\beta} \frac{\eta'(|y|)}{\eta(|y|)\,|y|}. \tag{6.55}$$

Similar to the symmetric case it is easy to see that the desired properties are fulfilled for the first argument.

It remains unclear what desirable properties for the other image are in this context.

▲

Example 6.2.15 (Quadratic Asymmetric Parallel Level Sets). The principal diffusivities of quadratic asymmetric parallel level sets are given by

$$\lambda_1^{\perp} = c(1 - |y_\eta|^2) \qquad \lambda_1^{\parallel} = c \tag{6.56}$$

$$\lambda_2^{\perp} = c\left(\langle x, y_\eta\rangle^2 \frac{\eta'(|y|)}{\eta(|y|)\,|y|} - \eta(|y|)^{-2}\,|x|^2\right) \qquad \lambda_2^{\parallel} = c\,\langle x, y_\eta\rangle^2 \frac{\eta'(|y|)}{\eta(|y|)\,|y|} \tag{6.57}$$

$$c \coloneqq \left(\beta^2 + |x|^2 - |\langle x, y_\eta\rangle|^2\right)^{-1/2}. \tag{6.58}$$

As in the previous cases, all the desirable properties hold true. In addition, notice that $\lambda_1^{\perp} \to 0$ for $|y| \to \infty$ as $1 - |y_\eta|^2$ can be seen as an edge indicator function. ▲

Example 6.2.16 (Kaipio). In Kaipio's model, i.e. $\varphi = id, \psi(s) = 1/2s^2, \beta = 0$, the principal diffusivities are

$$\lambda_1^{\perp} = 1 - |y_\eta|^2 \qquad \lambda_1^{\parallel} = 1 \tag{6.59}$$

$$\lambda_2^{\perp} = \langle x, y_\eta\rangle^2 \frac{\eta'(|y|)}{\eta(|y|)\,|y|} - \eta(|y|)^{-2}\,|x|^2 \qquad \lambda_2^{\parallel} = \langle x, y_\eta\rangle^2 \frac{\eta'(|y|)}{\eta(|y|)\,|y|}. \tag{6.60}$$

This diffusivities are similar to the quadratic model but without the edge-preserving properties for $y \to 0$. ▲

6.3. Discrete Gâteaux Derivative.

In this section we will discuss how to discretize the Gâteaux derivative of the priors presented so far. All these priors have a Gâteaux derivative of the form

$$\operatorname{div}(\kappa \nabla u) \tag{6.61}$$

with a scalar diffusivity κ or they consist of a sum of these terms. Therefore, for all priors the discretization of the Gâteaux derivative reduces to the discretization of (6.61).

We will first warm up with a one-dimensional example where we compute the discrete Gâteaux derivative when the gradient was discretized with forward differences. We will perform the same analysis for forward and central differences in two dimensions where we already introduce some notation that is needed for the arbitrary dimensional case. In addition, we present a common discretization of the Gâteaux derivative [20, 4] and compare it to the derivative of the discrete version. We complete this section by a

short discussion of the boundary conditions.

6.3.1. The One-Dimensional Case: Forward Differencing Scheme. Let $u \in \mathbb{R}^N$ be a sampling of an arbitrary function. As we do not yet want to consider boundary conditions let us assume for now that all terms are well-defined. We will discuss boundary conditions at the end of this section. Furthermore, let $f\colon \mathbb{R} \to \mathbb{R}$ be a differentiable function which defines the functional

$$\mathcal{P}(u) = \sum_{n=1}^{N} f(\partial^+ u_n) \tag{6.62}$$

and operates on forward finite differences

$$\partial^+ u_n \coloneqq u_{n+1} - u_n \tag{6.63}$$

of an image u at position n.

Then the partial derivative with respect to one sample u_m is given by

$$\partial_{u_m} f(\partial^+ u_n) = \partial_{u_m} f(u_{n+1} - u_n) \tag{6.64}$$

$$= f'(\partial^+ u_n)(\delta_{m,n+1} - \delta_{m,n}) = -f'(\partial^+ u_n)(\delta_{m,n} - \delta_{m-1,n}), \tag{6.65}$$

where we used the Kronecker delta $\delta_{n,m}$. To remind the reader, the Kronecker delta $\delta_{n,m}$ has been defined to be one, if $n = m$ and otherwise zero.

This at hand we can derive the partial derivative of $\mathcal{P}$ with respect to u_m.

$$D\mathcal{P}[u_m] = \partial_{u_m} \mathcal{P}(u) = \sum_{n=1}^{N} \{\partial_{u_m} f(u_{n+1} - u_n)\} \tag{6.66}$$

$$\overset{(6.65)}{=} -\sum_{n=1}^{N} \{f'(\partial^+ u_n)(\delta_{m,n} - \delta_{m-1,n})\} \tag{6.67}$$

$$= -[f'(\partial^+ u_m) - f'(\partial^+ u_{m-1})] = -\partial^- f'(\partial^+ u_m) \tag{6.68}$$

Here we have used backward finite differences

$$\partial^- u_n \coloneqq u_n - u_{n-1}. \tag{6.69}$$

Note also that for this notation to make sense, we have to identify $\{f'(\partial^+ u_m)\}$ as a vector that is indexed by m.

6.3.2. Finite Differences in Two Dimensions.

Forward Differencing Scheme. We will now consider the two-dimensional case, i.e. $u \in \mathbb{R}^N \times \mathbb{R}^N$, and generalize the previous result. We define

$$\partial_1^+ u_{n,m} \coloneqq u_{n+1,m} - u_{n,m}, \qquad \partial_2^+ u_{n,m} \coloneqq u_{n,m+1} - u_{n,m} \tag{6.70}$$

and the discrete gradient $[\nabla^+ u_{n,m}]_d \coloneqq \partial_d^+ u_{n,m}$. It is convenient to use a generic expression for both partial derivatives. Let $\alpha = (n, m)$ be a multi-index and $(e_d)_k \coloneqq \delta_{d,k}$ the dth standard basis vector. Then there is

$$\partial_d^+ u_{n,m} \coloneqq u_{\alpha+e_d} - u_\alpha. \tag{6.71}$$

We have all the notation at hand to compute the two-dimensional discrete Gâteaux derivative. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a differentiable function which again defines the functional $\mathcal{P}$ as

$$\mathcal{P}(u) = \sum_{\alpha \in \{1,\dots,N\}^2} f(\nabla^+ u_\alpha) = \sum_{n=1}^{N} \sum_{m=1}^{N} f(\nabla^+ u_{n,m}). \tag{6.72}$$

In the same manner as in the one-dimensional case we can compute the partial derivatives of f with respect to u_α. First, observe that

$$\partial_{u_\alpha} \partial_d^+ u_\beta = \partial_{u_\alpha}(u_{\beta+e_d} - u_\beta) = \delta_{\alpha,\beta+e_d} - \delta_{\alpha,\beta} = -\big(\delta_{\alpha,\beta} - \delta_{\alpha-e_d,\beta}\big) \tag{6.73}$$

and therefore

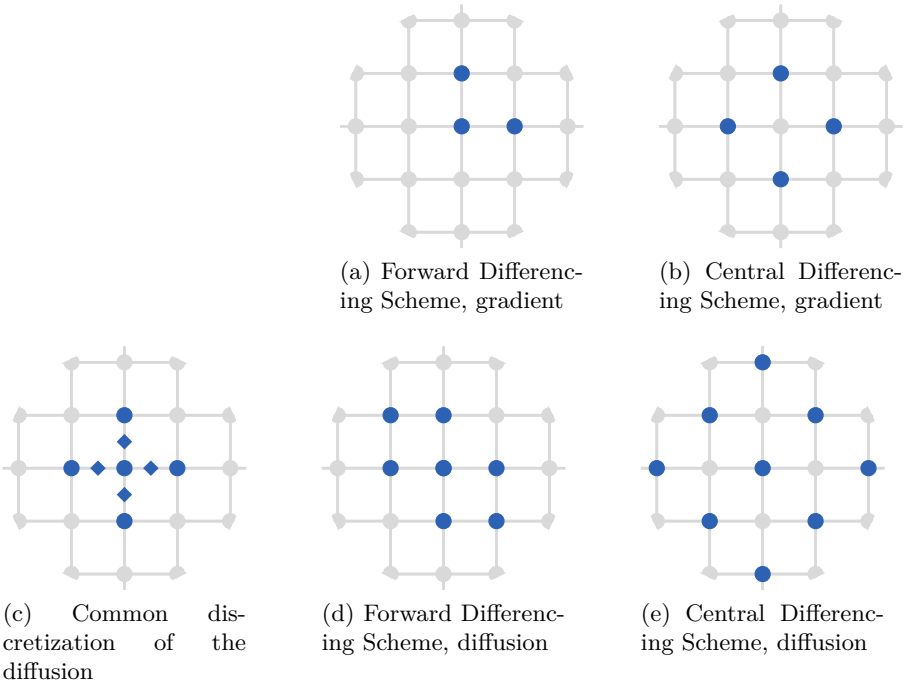
$$\partial_{u_\alpha} f(\nabla^+ u_\beta) = \sum_{d=1}^{2} \partial_d f(\nabla^+ u_\beta) \partial_{u_\alpha} \partial_d^+ u_\beta \overset{(6.73)}{=} -\sum_{d=1}^{2} \partial_d f(\nabla^+ u_\beta)\big(\delta_{\alpha,\beta} - \delta_{\alpha-e_d,\beta}\big). \tag{6.74}$$

Then the Gâteaux derivative of $\mathcal{P}$ at u_α is given by

$$D\mathcal{P}[u_\alpha] = \sum_{\beta \in \{1,\dots,N\}^2} \partial_{u_\alpha} f(\nabla^+ u_\beta) \overset{(6.74)}{=} -\sum_{d=1}^{2} \sum_{\beta \in \{1,\dots,N\}^2} \partial_d f(\nabla^+ u_\beta)\big(\delta_{\alpha,\beta} - \delta_{\alpha-e_d,\beta}\big) \tag{6.75}$$

$$= -\sum_{d=1}^{2} \big\{\partial_d f(\nabla^+ u_\alpha) - \partial_d f(\nabla^+ u_{\alpha-e_d})\big\} \tag{6.76}$$

$$= -\sum_{d=1}^{2} \partial_d^- \partial_d f(\nabla^+ u_\alpha) = -\operatorname{div}^-\big[\nabla f(\nabla^+ u_\alpha)\big]. \tag{6.77}$$



(a) Forward Differencing Scheme, gradient

(b) Central Differencing Scheme, gradient

(c) Common discretization of the diffusion

(d) Forward Differencing Scheme, diffusion

(e) Central Differencing Scheme, diffusion

Figure 6.5. *Marked are the pixels used to compute the gradient ∇u (top) and diffusion $\operatorname{div}(\kappa\nabla u)$(bottom). In the standard approach (c) inner terms are approximated on a subgrid (diamonds). This leads to a small filter width and a symmetric update scheme. If the gradient is approximated using forward (a) or central differences (b), the diffusion term has a unique discretization that matches the discrete Gâteaux derivative. While the central differencing scheme (e) uncouples half of the pixels and leads to checker board like artefacts, the forward differencing scheme (d) has a small but asymmetric support.*

Here we used the discrete backward divergence operator

$$\operatorname{div}^- v_\alpha = \sum_{d=1}^{2} (\partial_d^- v_d)_\alpha. \tag{6.78}$$

So, we see that if the gradient is approximated with forward differences, then the divergence has to be discretized with backward differences so that the discretized derivative is the derivative of the discretized functional.

We see from (6.76) that for the derivative at (n, m) we need to compute

$$\nabla^+ u_{n,m}, \quad \nabla^+ u_{n-1,m}, \quad \text{and} \quad \nabla^+ u_{n,m-1} \tag{6.79}$$

which involves u sampled at seven locations, see figure 6.5.

Central Differencing Scheme. Consider now the two-dimensional case where the gradient is not approximated with forward but with central finite differences

$$\partial_d^c u_\alpha = \tfrac{1}{2}(u_{\alpha+e_d} - u_{\alpha-e_d}). \tag{6.80}$$

so we want to compute the derivative of

$$\mathcal{P}(u) = \sum_{\alpha\in\{1,\dots,N\}^2} f(\nabla^c u_\alpha). \tag{6.81}$$

Then there is

$$\partial_{u_\alpha}\partial_d^c u_\beta = \tfrac{1}{2}\partial_{u_\alpha}(u_{\beta+e_d} - u_{\beta-e_d}) \tag{6.82}$$

$$= \tfrac{1}{2}(\delta_{\alpha,\beta+e_d} - \delta_{\alpha,\beta-e_d}) = -\tfrac{1}{2}(\delta_{\alpha+e_d,\beta} - \delta_{\alpha-e_d,\beta}) \tag{6.83}$$

and hence

$$\partial_{u_\alpha} f(\nabla^c u_\beta) = \sum_{d=1}^{2} \partial_d f(\nabla^c u_\beta)\partial_{u_\alpha}\partial_d^c u_\beta = -\sum_{d=1}^{2} \tfrac{1}{2}\partial_d f(\nabla^c u_\beta)(\delta_{\alpha+e_d,\beta} - \delta_{\alpha-e_d,\beta}). \tag{6.84}$$

We derive the Gâteaux derivative of $\mathcal{P}$ at u_α as

$$D\mathcal{P}[u_\alpha] = \sum_{\beta\in\{1,\dots,N\}^2} \partial_{u_\alpha} f(\nabla^c u_\beta) \tag{6.85}$$

$$= -\sum_{d=1}^{2} \tfrac{1}{2} \sum_{\beta\in\{1,\dots,N\}^2} \partial_d f(\nabla^c u_\beta)(\delta_{\alpha+e_d,\beta} - \delta_{\alpha-e_d,\beta}) \tag{6.86}$$

$$= -\sum_{d=1}^{2} \tfrac{1}{2}\{\partial_d f(\nabla^c u_{\alpha+e_d}) - \partial_d f(\nabla^c u_{\alpha-e_d})\} \tag{6.87}$$

$$= -\sum_{d=1}^{2} \partial_d^c \partial_d f(\nabla^c u_\alpha) = -\operatorname{div}^c[\nabla f(\nabla^c u_\alpha)] \tag{6.88}$$

with the discrete central divergence operator

$$\operatorname{div}^c v_\alpha = \sum_{d=1}^{2} (\partial_d^c v_d)_\alpha. \tag{6.89}$$

This means that the matching discretization to central differences for the gradient are central differences for the divergence.

As in the case of forward differences, we want to see which samples of u are involved in the computation of the derivative at (n, m). From (6.87) we see that we need to

compute

$$\nabla^c u_{n+1,m}, \quad \nabla^c u_{n,m+1}, \quad \nabla^c u_{n-1,m}, \quad \nabla^c u_{n,m-1} \tag{6.90}$$

which involves in total nine samples of u, see figure 6.5.

6.3.3. Arbitrary Differencing Scheme in any Dimensions. From the previous examples in one and two dimensions with forward and central differences, we have seen that every time very similar computations were needed to derive the discrete Gâteaux derivative. We will state now the general result for arbitrary sampling in arbitrary dimensions.

Theorem 6.3.1 (Derivative of a Discrete Functional). *Let* $\mathsf{U} \coloneqq \{1, \dots, N\}^D$ *be the set of all sample locations and* $\mathsf{X} \coloneqq \{u \colon \mathsf{U} \to \mathbb{R}\}$ *the space of all discrete images sampled at these locations. Moreover, let* $f \in \mathsf{C}^1(\mathbb{R}^D), \gamma \in \mathbb{R}^{D \times K}, \xi \in \mathbb{Z}^{D \times K}$ *and*

$$A_d \colon \mathsf{X} \to \mathsf{X}, \quad (A_d u)_\alpha \coloneqq \sum_{k=1}^{K} \gamma_{d,k} u_{\alpha + \xi_{d,k}} \tag{6.91}$$

$$B_d \colon \mathsf{X} \to \mathsf{X}, \quad (A_d u)_\alpha \coloneqq \sum_{k=1}^{K} \gamma_{d,k} u_{\alpha - \xi_{d,k}} \tag{6.92}$$

be generalized finite differences. Furthermore, let $[(Au)_\alpha]_d \coloneqq (A_d u)_\alpha$ *be a generalized gradient approximation and* $(Bu)_\alpha \coloneqq \sum_{d=1}^{D} (B_d u)_\alpha$ *a generalized divergence approximation.*

Then the derivative of the discrete functional

$$\mathcal{P} \colon \mathsf{X} \to \mathbb{R}, \quad u \mapsto \sum_{\alpha \in \mathsf{U}} f(A u_\alpha). \tag{6.93}$$

is given by

$$D\mathcal{P}[u_\alpha] = [B \nabla f(Au)]_\alpha. \tag{6.94}$$

Before we proof this theorem let us check that the previous examples are indeed special cases of this theorem.

Example 6.3.2 (Forward Differences). Let us consider forward differences in any dimension. Then γ does not depend on the dimension, i.e. $\gamma_{d,k} = \gamma_k$ and ξ can be decomposed as $\xi_{d,k} = \xi_k e_d$. Then we recover forward differences by the choice $\gamma = (1, -1)$

and $(\xi_k)_k = (1, 0)$

$$(A_d u)_\alpha = u_{\alpha+e_d} - u_\alpha = \partial_d^+ u_\alpha. \tag{6.95}$$

With this choice the operator B_m are backward differences

$$(B_d u)_\alpha = u_{\alpha-e_d} - u_\alpha = -(u_\alpha - u_{\alpha-e_d}) = -\partial_d^- u_\alpha. \tag{6.96}$$

▲

Example 6.3.3 (Central Differences). Let us now turn to central differences in any dimension. Again as for the forward differences γ is dimension independent, i.e. $\gamma_{d,k} = \gamma_k$, and ξ can be decomposed as $\xi_{d,k} = \xi_k e_d$. For central differences, we need to choose $\gamma = (1/2, -1/2)$ and $(\xi_k) = (1, -1)$. Then there is

$$(A_d u)_\alpha = \tfrac{1}{2} u_{\alpha+e_d} - \tfrac{1}{2} u_{\alpha-e_d} = \tfrac{1}{2}(u_{\alpha+e_d} - u_{\alpha-e_d}) = \partial_d^c u_\alpha \tag{6.97}$$

and

$$(B_d u)_\alpha = \tfrac{1}{2} u_{\alpha-e_d} - \tfrac{1}{2} u_{\alpha+e_d} = -\tfrac{1}{2}(u_{\alpha+e_d} - u_{\alpha-e_d}) = -\partial_d^c u_\alpha. \tag{6.98}$$

▲

Proof of theorem 6.3.1. In the same manner as in the one- and two-dimensional cases, we compute the partial derivatives of f with respect to u_α. First, observe that

$$\partial_{u_\alpha}(A_d u)_\beta = \sum_{k=1}^{K} \gamma_{d,k} \partial_{u_\alpha} u_{\beta+\xi_{d,k}} = \sum_{k=1}^{K} \gamma_{d,k} \delta_{\alpha,\beta+\xi_{d,k}} = \sum_{k=1}^{K} \gamma_{d,k} \delta_{\alpha-\xi_{d,k},\beta} \tag{6.99}$$

and therefore

$$\partial_{u_\alpha} f(Au_\beta) = \sum_{d=1}^{D} \partial_d f(Au_\beta) \partial_{u_\alpha}(A_d u)_\beta \overset{(6.99)}{=} \sum_{d=1}^{D} \sum_{k=1}^{K} \gamma_{d,k} \partial_d f(Au_\beta) \delta_{\alpha-\xi_{d,k},\beta}. \tag{6.100}$$

Finally, the Gâteaux derivative of $\mathcal{P}$ at u_α is then

$$D\mathcal{P}[u_\alpha] = \sum_{\beta \in \mathsf{U}} \partial_{u_\alpha} f(Au_\beta) \overset{(6.100)}{=} \sum_{d=1}^{D} \sum_{k=1}^{K} \gamma_{d,k} \sum_{\beta \in \mathsf{U}} \partial_d f(Au_\beta) \delta_{\alpha-\xi_{d,k},\beta} \tag{6.101}$$

$$= \sum_{d=1}^{D} \sum_{k=1}^{K} \gamma_{d,k} \partial_d f(Au_{\alpha-\xi_{d,k}}) = \sum_{d=1}^{D} B_d \partial_d f(Au_\alpha) = [B \nabla f(Au)]_\alpha. \tag{6.102}$$

■

6.3.4. Discretization of the Diffusion Term. So far we were concerned about a discretization of the Gâteaux derivative that matches the derivative of the discrete functional. Therefore, we do not have much control about symmetry and the size of the support, cf. figure 6.5. A common strategy is a different approach which discretizes the Gâteaux derivative directly. We will make use of central finite differences on a subgrid

$$\partial_1^c u_{n,m} \coloneqq u_{n+1/2,m} - u_{n-1/2,m} \tag{6.103}$$

which for sake of simplicity we denote by the same symbol as "usual" central differences. As the values on the subgrid are not defined yet, we interpolate them linearly from the original values

$$u_{n+1/2,m} \coloneqq \tfrac{1}{2}(u_{n+1,m} + u_{n,m}). \tag{6.104}$$

The trick is that we take twice central differences on a subgrid so that we do not need to interpolate many values. The discretization and interpolation in the second direction are defined similarly.

Let us next calculate $\partial_1^c(\kappa_{n,m}\partial_1^c u_{n,m})$. This is given by

$$\begin{aligned}
\partial_1^c(\kappa_{n,m}\partial_1^c u_{n,m}) &= \partial_1^c\big[\kappa_{n,m}(u_{n+1/2,m} - u_{n-1/2,m})\big] && (6.105)\\
&= \kappa_{n+1/2,m}(u_{n+1,m} - u_{n,m}) - \kappa_{n-1/2,m}(u_{n,m} - u_{n-1,m}) && (6.106)\\
&= \tfrac{1}{2}(\kappa_{n+1,m} + \kappa_{n,m})(u_{n+1,m} - u_{n,m})\\
&\quad - \tfrac{1}{2}(\kappa_{n,m} + \kappa_{n-1,m})(u_{n,m} - u_{n-1,m}). && (6.107)
\end{aligned}$$

As the same holds true for the second direction, we derive

$$\begin{aligned}
&\operatorname{div}(\kappa_{n,m}\nabla u_{n,m}) && (6.108)\\
&\quad \approx \partial_1^c(\kappa_{n,m}\partial_1^c u_{n,m}) + \partial_2^c(\kappa_{n,m}\partial_2^c u_{n,m}) && (6.109)\\
&\overset{(6.107)}{=} \tfrac{1}{2}(\kappa_{n+1,m} + \kappa_{n,m})(u_{n+1,m} - u_{n,m})\\
&\qquad - \tfrac{1}{2}(\kappa_{n,m} + \kappa_{n-1,m})(u_{n,m} - u_{n-1,m})\\
&\qquad + \tfrac{1}{2}(\kappa_{n,m+1} + \kappa_{n,m})(u_{n,m+1} - u_{n,m})\\
&\qquad - \tfrac{1}{2}(\kappa_{n,m} + \kappa_{n,m-1})(u_{n,m} - u_{n,m-1}) && (6.110)
\end{aligned}$$

$$= \sum_{(k,l)\in \mathsf{N}(n,m)} \tfrac{1}{2}(\kappa_{k,l} + \kappa_{n,m}) u_{k,l} - \left\{ \sum_{(k,l)\in \mathsf{N}(n,m)} \tfrac{1}{2}(\kappa_{k,l} + \kappa_{n,m}) \right\} u_{n,m}. \qquad (6.111)$$

Here, we used the "neighbourhood" $\mathsf{N}(n,m) := \{(n+1,m),(n,m+1),(n-1,m),(n,m-1)\}$.

Compared to the previous approaches, we can see from figure 6.5 that the support is now small and symmetric. In addition, (6.111) shows that not only is the support symmetric but so is the whole derivative.

On the downside, there is probably no discretized functional so that this is a matching derivative. Therefore it is not recommended to use this discretization in gradient based optimization methods.

6.3.5. Boundary Conditions. In the discussion above we intentionally neglected boundary conditions but in practice we always consider a bounded domain so we need to deal with boundary conditions. In theorem 4.2.5 one of the assumptions are Neumann boundary conditions, i.e. vanishing normal derivatives. This condition ensures that there is no flow out of the image, hence the prior/regularization preserves the energy of the image.

To ensure Neumann boundary conditions for

$$\operatorname{div}^{-}\left[\nabla f(\nabla^{+} u)\right] \qquad (6.112)$$

we extend the image u for the forward derivatives constantly and for the backward differences we extend $\nabla f(\nabla^{+} u)$ with zeros. As the backward differences in the divergence operate on the partial derivatives this ensures that the missing partial derivatives are treated as zero.

6.4. Summary.

In this chapter we have computed the Gâteaux derivative of both symmetric and asymmetric parallel level sets. It turned out that the corresponding diffusive flow is anisotropic in all cases with principal directions given by the local coordinates defined by the other image. We discussed several ways of discretizing the Gâteaux derivative. Finite forward differences for the gradient and the matching backward differences for the divergence are a good compromise as they provide a matching functional-derivative pair and have a small support.

Chapter 7

Joint Reconstruction and Convexity

Convexity is an important property in optimization as it allows to find global minimizers with local gradient information. The main aim of this chapter is to get an understanding of convexity in a joint reconstruction set up and to analyse the convexity of the proposed priors.

7.1. Convex Sets and Convex Functions.

Definition 7.1.1 (Convex Set). *Let* $\mathsf{U} \subset \mathsf{X}$ *be a subset of a vector space. We call* U convex *if for all* $x_1, x_2 \in \mathsf{X}$ *the line segment joining them is inside* U*, i.e. for all* $s \in [0,1]$ *there is* $sx_1 + (1-s)x_2 \in \mathsf{U}$.

Definition 7.1.2 (Convex Function). *Let* $\mathsf{U} \subset \mathsf{X}$ *be a convex set. Then a function* $f\colon \mathsf{U} \to \mathbb{R}$ *is said to be* convex *if for all* $s \in [0,1], x_1, x_2 \in \mathsf{U}$ *there is*

$$f(sx_1 + (1-s)x_2) \leq sf(x_1) + (1-s)f(x_2). \tag{7.1}$$

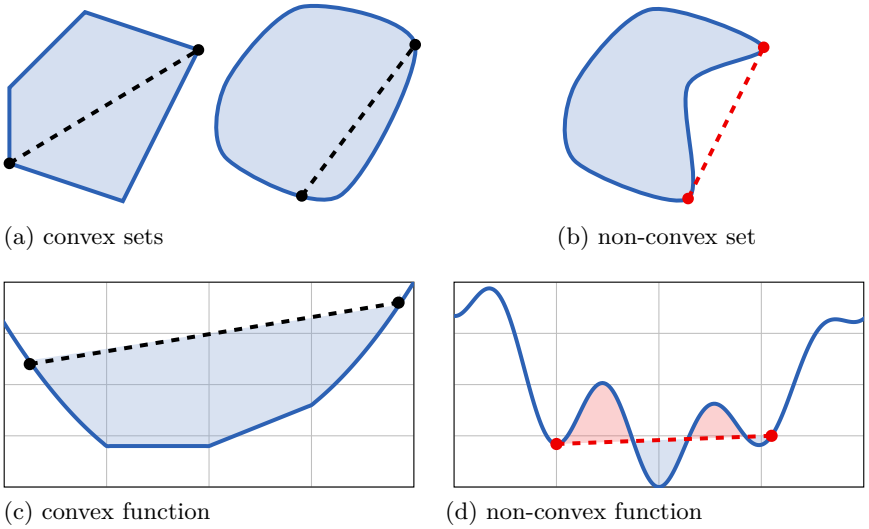
A few examples of convex / non-convex sets and functions are presented in figure 7.1.

Remark 7.1.3 (Multi-Dimensional Convexity). From the definition of convexity we see that a function $f\colon \mathsf{U} \to \mathbb{R}$ is convex if and only if for all $x_1, x_2 \in \mathsf{U}$ the function

$$g\colon [0,1] \to \mathbb{R}, \qquad g(s) = f(x_1 + s(x_2 - x_1)) \tag{7.2}$$

is convex [116, p. 114]. This means that a function is convex if and only if it, restricted to all possible combinations of one-dimensional segments, is convex. ▲

Remark 7.1.4 (Geometric Characterization of Convexity). The definition of convexity is algebraic but there is an equivalent geometric definition as well. For a function $f\colon \mathsf{U} \to \mathbb{R}$ and $s \in \mathbb{R}$ we call $\{x \in \mathsf{X} \mid f(x) \leq s\}$ a *sub-level set* of f. A function f is convex if



(a) convex sets (b) non-convex set

(c) convex function (d) non-convex function

Figure 7.1. *Examples for convex and non-convex sets and functions. For convex sets the line segment between any pair of points in the set must lie in the set as well. This is the case in a) but not for b). Similarly, the graph of a convex function must lie underneath the line segment between two points on the graph which is the case in c) but not in d).*

and only if all of its sub-level sets are convex [116, p. 116]. This characterization gives us a visual indication if a function is convex or not. ▲

Lemma 7.1.5 (Basic Operations with Convexity). *Let* $\mathsf{U} \subset \mathsf{X}$ *be a convex subset,* X, Y *be vector spaces and* $f\colon \mathsf{U} \to \mathbb{R}$ *be convex. Moreover, let* $s \geq 0$, $A \in \mathsf{L}(\mathsf{X}, \mathsf{Y})$, $g\colon \mathsf{U} \to \mathbb{R}$ *be convex and* $\Phi\colon f(\mathsf{U}) \to \mathbb{R}$ *be monotonically increasing and convex.*

Then, all of the functions i) tf*, ii)* $A|_{A^{-1}(\mathsf{U})} \circ f$*, iii)* $f + g$*, iv)* $f \circ \Phi$ *are convex.*

Proof. Let $x_1, x_2 \in \mathsf{U}, s \in [0, 1]$. Then there is

$$
\begin{aligned}
(tf)(A(sx_1 + (1-s)x_2)) &= tf(sAx_1 + (1-s)Ax_2) && (7.3)\\
&\leq t\big[sf(Ax_1) + (1-s)f(Ax_2)\big] && (7.4)\\
&= s(tf)(Ax_1) + (1-s)(tf)(Ax_2) && (7.5)
\end{aligned}
$$

which shows i) and ii). Similarly,

$$
\begin{aligned}
(f+g)(sx_1 + (1-s)x_2) &= f(sx_1 + (1-s)x_2) + g(sx_1 + (1-s)x_2) && (7.6)\\
&\leq sf(x_1) + (1-s)f(x_2) + sg(x_1) + (1-s)g(x_2) && (7.7)\\
&= s(f+g)(x_1) + (1-s)(f+g)(x_2) && (7.8)
\end{aligned}
$$

proves iii). Finally, there is

$$\Phi(f(sx_1 + (1-s)x_2)) \leq \Phi(sf(x_1) + (1-s)f(x_2)) \tag{7.9}$$

$$\leq s\Phi(f(x_1)) + (1-s)\Phi(f(x_2)). \tag{7.10}$$

Here we first used the monotonicity of Φ and the convexity of f and then the convexity of Φ. Thus, iv) is proven. ■

Definition 7.1.6 (Positive-Semidefinite Matrices). *A matrix $A \in \mathbb{R}^{D\times D}$ is said to be* positive-semidefinite *if for all $x \in \mathbb{R}^D$ there is $x^T Ax \geq 0$.*

Proposition 7.1.7 (Convexity and Smoothness, [26, p. 195]). *Let $\mathsf{U} \subset \mathbb{R}^D$ be an convex, open set and $f\colon \mathsf{U} \to \mathbb{R}$ be twice continuously differentiable. Then f is convex if and only if its Hessian is positive-semidefinite everywhere.*

7.2. Convexity for Gradient based Regularization.

In this section we want to characterize the convexity of functionals of the form

$$\mathcal{P}(u) = \int_\Omega f(\nabla u(x))\,\mathrm{d}x. \tag{7.11}$$

Under some assumptions we will show that $\mathcal{P}$ is convex if and only if f is convex. This shall be made more precise and proven in the following.

Theorem 7.2.1. *Let $\Omega \subset \mathbb{R}^D$ be a set of non-zero and finite "size" $|\Omega| \coloneqq \int_\Omega 1\,\mathrm{d}x$, i.e. $0 < |\Omega| < \infty$. Let $A\colon \mathsf{X} \to \mathsf{Y}$ be a linear operator between two vector spaces $\mathsf{X} \subset \{u\colon \Omega \to \mathbb{R}^M\}, \mathsf{Y} \coloneqq \{u\colon \Omega \to \mathbb{R}^K\}$. Moreover, let $f\colon \mathbb{R}^K \to [0,\infty)$ and $\mathsf{U} \subset \mathsf{X}$ be a convex set such that $\int_\Omega f(Au(x))\,\mathrm{d}x < \infty$ for all $u \in \mathsf{U}$ and $\{u\colon \Omega \to \mathbb{R}^M \mid u \equiv c \in \mathbb{R}^M\} \subset A(\mathsf{U})$.*

Then, the functional

$$\mathcal{P}\colon \mathsf{U} \to \mathbb{R}, \qquad \mathcal{P}(u) = \int_\Omega f(Au(x))\,\mathrm{d}x \tag{7.12}$$

is convex if and only if f is convex.

Corollary 7.2.2. *In addition to the assumptions of theorem 7.2.1, let $\Omega \subset \mathbb{R}^D$ be open, $\mathsf{X} \subset \mathsf{C}^1(\Omega, \mathbb{R}^M)$ and $A = \nabla$ be a "gradient operator" of vector-valued images $u\colon \Omega \to \mathbb{R}^M$, i.e. $\nabla\colon \mathsf{X} \to \{u\colon \Omega \to \mathbb{R}^{D\cdot M}\}$ with $(\nabla u)_{(m-1)D+d} \coloneqq (\nabla u_m)_d$.*

Then $\mathcal{P}$ given by (7.11) is convex if and only if f is convex.

Remark 7.2.3. It is likely that the assumptions of corollary 7.2.2 can be relaxed. For example the strong condition $\{u\colon \Omega \to \mathbb{R}^M \mid u \equiv c \in \mathbb{R}^M\} \subset \nabla(\mathsf{U})$ prevents arbitrary boundary conditions. For instance, for zero boundary conditions the inclusion can only be fulfilled for $c = 0$. Probably this can be relaxed using mollifiers [81, p. 260] which would have made the proof a lot more technical. ▲

Proof of theorem 7.2.1. Assume that f is convex and let $u_1, u_2 \in \mathsf{U}$, $s \in [0,1]$. From the convexity of f and basic integral properties, it follows that

$$\begin{aligned}\mathcal{P}(su_1 + (1-s)u_2) &= \int_\Omega f(sAu_1(x) + (1-s)Au_2(x))\,\mathrm{d}x && (7.13)\\ &\leq \int_\Omega sf(Au_1(x)) + (1-s)f(Au_2(x))\,\mathrm{d}x && (7.14)\\ &= s\mathcal{P}(u_1) + (1-s)\mathcal{P}(u_2). && (7.15)\end{aligned}$$

Assume now that f is not convex. Then there exist vectors $y_1, y_2 \in \mathbb{R}^K$ and $s \in (0,1)$ so that

$$f(sy_1 + (1-s)y_2) > sf(y_1) + (1-s)f(y_2). \tag{7.16}$$

By assumption, there are $u_1, u_2 \in \mathsf{U}$ such that $Au_k \equiv y_k$ for $k \in \{1,2\}$. Hence,

$$\begin{aligned}\mathcal{P}(su_1 + (1-s)u_2) &= \int_\Omega f(sy_1 + (1-s)y_2)\,\mathrm{d}x = |\Omega| f(sy_1 + (1-s)y_2) && (7.17)\\ &\overset{(7.16)}{>} |\Omega|\big[sf(y_1) + (1-s)f(y_2)\big] = s\mathcal{P}(u_1) + (1-s)\mathcal{P}(u_2). && (7.18)\end{aligned}$$

■

Remark 7.2.4 (Invexity). Optimization problems can be divided into the classes of convex and non-convex problems. It is sometimes said that these classes divide the optimization problems into "simple" and "non-simple" problems. It is shown in figure 7.2 that some non-convex problems can be easily minimized by gradient type methods. These functions are coercive, smooth and have a unique minimum in zero but are not convex.

Functions for which a vanishing gradient at a minimum is a necessary and sufficient condition are called *invex* [109]. Invex functions have the undesirable property that sums of invex functions are not necessarily invex any more. ▲

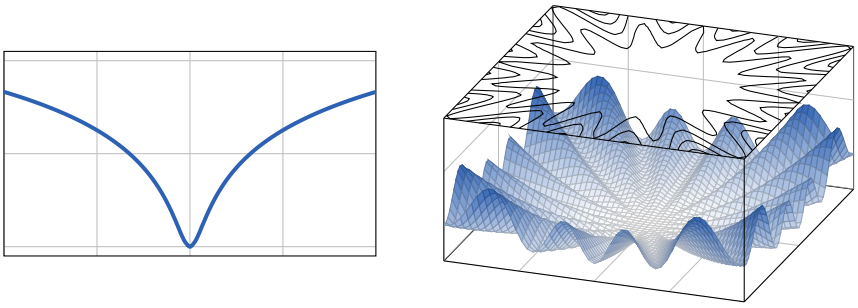


Figure 7.2. *Example of non-convex functions that can be optimized with gradient based optimization. The figure shows the graph of the functions* $f_1(x) = \log(x^2 + 1)$ *(left) and* $f_2(x, y) = [2 + \cos(4\,\mathrm{atan2}(y, x))] \cdot (x^2 + y^2)$ *(right). They are smooth, coercive and have a unique minimum at zero but they are not convex. Obviously, these are contrived examples but this shows that convexity is not necessary for gradient based global optimization.*

7.3. The Necessary Growth Condition and its Implications.

In this section we investigate necessary conditions for convexity that can be used to easily show that a function is not convex. In particular, we will take the intrinsic multi-dimensional structure of a joint prior into account.

Theorem 7.3.1 (Necessary Condition for Multi-Dimensional Convexity). *Let* $\mathsf{U} \subset \mathsf{X} \times \mathsf{X}$ *be a convex set and* $f \colon \mathsf{U} \to \mathbb{R}$ *a convex function. Then for all* $(x, y) \in \mathsf{U}$ *there holds the* growth condition

$$2\, f(\tfrac{1}{2}x, \tfrac{1}{2}y) \leq f(x, 0) + f(0, y). \tag{7.19}$$

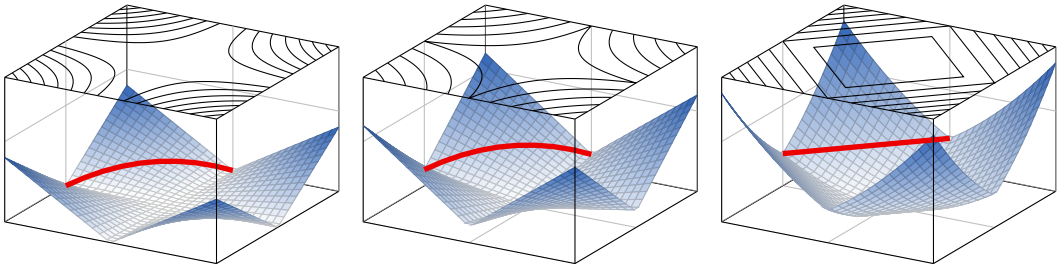
Proof. Let $s = \frac{1}{2}, z_0 = (x, 1)$ and $z_2 = (0, y)$. Then the claim follows immediately from the definition of convexity. ∎

Corollary 7.3.2. *Let* $f \colon \mathsf{X} \times \mathsf{X} \to \mathbb{R}$ *be a convex function that can be decomposed as* $f(x, y) = g(x) + g(y) + h(x, y)$ *with* $g(0) = 0$ *and* $h(x, 0) = 0$. *Then for all* $x \in \mathsf{X}$ *there is*

$$h(\tfrac{1}{2}x, \tfrac{1}{2}x) \leq g(x) - 2g(\tfrac{1}{2}x). \tag{7.20}$$

Example 7.3.3 (p-homogeneous functions). Let f be non-negative, convex and p-homogeneous, i.e. for all $s \geq 0$ and $x \in \mathsf{X}$ there is $f(sx) = s^p f(x)$. Then, there is $p \geq 1$. ▲

Proof. Corollary 7.3.2 implies $0 \leq f(x)(1 - 2^{1-p})$ which in turn yields $2^{1-p} \leq 1$ with the non-negativity of f. ∎



(a) f is not convex (b) $f + g_1$ is not convex (c) $f + g_2$ is convex

Figure 7.3. *Illustration of example 7.3.5. The figure shows the function $f(x, y) = |x|\,|y|$ with two different convex functions functions $g_1(x, y) = \frac{1}{2}(|x| + |y|)$ and $g_2(x, y) = \frac{1}{2}(|x|^2 + |y|^2)$ being added. While a) f and b) $f + g_1$ are non-convex, the addition of g_2 makes c) $f + g_2$ a convex function. Note that $(f + g_2)(x, y) = \frac{1}{2}(|x| + |y|)^2$.*

Remark 7.3.4. The example shows a necessary condition for convexity of the parallel level sets functional. If $\psi \circ \varphi$ is p-homogeneous, then $p \geq 1$ is necessary (but not sufficient) for convexity in each argument. ▲

Example 7.3.5. Let $f(x, y) = |x|^q\,|y|^q + c\,|x|^p + c\,|y|^p$ be convex. Then $p = 2q$ and $c \geq (2^p - 2)^{-1}$. This is highlighted in figure 7.3. ▲

Proof. Corollary 7.3.2 yields $2^{-2q}\,|x|^{2q} \leq (1 - 2^{1-p})c\,|x|^p$ and in particular for all $x \neq 0$

$$|x|^{2q-p} \leq 2^{2q}(1 - 2^{1-p})c. \tag{7.21}$$

The boundedness for small and large $|x|$ implies $p = 2q$ which in turn implies $c \geq (2^p - 2)^{-1}$. ■

7.4. Convexity of Joint Prior Examples.

In this section we will have a look at the convexity of the priors for multi-modality imaging in the last chapters.

Example 7.4.1 (Separate Total Variation). Separate total variation, cf. (4.24), is convex in each argument as well as in the joint argument. ▲

Proof. It is well known that total variation for scalar-valued functions is convex. Thus, it follows that separate total variation is convex as a sum of convex functions, cf. lemma 7.1.5. ■

Example 7.4.2 (Joint Total Variation). Joint total variation, cf. (4.28), is convex in each argument as well as in the joint argument. ▲

Proof. Let $z = (x, y)$. Then the regularization function of joint total variation is $f_{\mathbb{TV}^{\mathcal{J}}}(x, y) = \left(|x|^2 + |y|^2\right)^{1/2} = |z|$, hence convex in the joint argument.

The convexity in each argument follows from the convexity in the joint argument after remark 7.1.4. ∎

Example 7.4.3 (Cross-Gradients). While the cross-gradients functional, cf. (5.26), is convex in each argument it is not convex in the joint argument. ▲

Proof. We will show that the Hessian of

$$f_{\mathrm{CG}}(x, y) = \frac{1}{2}\left(|x|^2 |y|^2 - \langle x, y\rangle^2\right) \tag{7.22}$$

with respect to one argument is positive-semidefinite but the Hessian with respect to the joint argument is not.

Straightforward computations yield the gradient and Hessian of f_{CG} with respect to x.

$$\partial_{x_n} f_{\mathrm{CG}}(x, y) = x_n |y|^2 - \langle x, y\rangle y_n \tag{7.23}$$

$$\partial_{x_m, x_n} f_{\mathrm{CG}}(x, y) = \delta_{n,m} |y|^2 - y_n y_m \tag{7.24}$$

Then there is for any $z \in \mathbb{R}^N$

$$z^T H f_{\mathrm{CG}}(x, y) z = z^T \{|y|^2 I - yy^T\} z = |y|^2 |z|^2 - \langle y, z\rangle^2 \geq 0 \tag{7.25}$$

and hence f_{CG} is convex in the first argument. As the functional is symmetric in its arguments this property also holds for the second argument.

Next, we will show that f_{CG} is not convex in the joint argument (x, y). The mixed second derivatives are given by

$$\partial_{y_m, x_n} f_{\mathrm{CG}}(x, y) = 2x_n y_m - y_n x_m - \langle x, y\rangle \delta_{n,m}. \tag{7.26}$$

Let $z_1 = y, z_2 = -x$ and x, y both non-zero and not parallel. Then there is

$$[z_1; z_2]^T H f_{\mathrm{CG}}(x, y)[z_1; z_2] \tag{7.27}$$

$$= [z_1; z_2]^T \begin{pmatrix} |y|^2 I - yy^T & 2xy^T - yx^T - \langle x, y\rangle I \\ 2xy^T - yx^T - \langle x, y\rangle I & |x|^2 I - xx^T \end{pmatrix} [z_1; z_2] \tag{7.28}$$

$$
\begin{aligned}
&= |y|^2 |z_1|^2 - \langle y, z_1\rangle^2 + 2\langle x, z_1\rangle \langle y, z_2\rangle - \langle x, z_2\rangle \langle y, z_1\rangle - \langle x, y\rangle \langle z_1, z_2\rangle \\
&\quad + 2\langle x, z_2\rangle \langle y, z_1\rangle - \langle x, z_1\rangle \langle y, z_2\rangle - \langle x, y\rangle \langle z_1, z_2\rangle + |x|^2 |z_2|^2 - \langle x, z_2\rangle^2 \quad (7.29) \\
&= |y|^2 |z_1|^2 - \langle y, z_1\rangle^2 + |x|^2 |z_2|^2 - \langle x, z_2\rangle^2 \\
&\quad + \langle x, z_1\rangle \langle y, z_2\rangle - 2\langle x, y\rangle \langle z_1, z_2\rangle + \langle x, z_2\rangle \langle y, z_1\rangle \quad (7.30) \\
&= -(|x|^2 |y|^2 - \langle x, y\rangle^2) < 0. \quad (7.31)
\end{aligned}
$$

■

Motivated by example 7.3.5 and figure 7.3 we can "convexify" the cross-gradients functional.

Lemma 7.4.4. *The modified cross-gradients functional*

$$f(x, y) = \frac{1}{2}(|x|^4 + |y|^4 + |x|^2 |y|^2 - \langle x, y\rangle^2) \quad (7.32)$$

is convex in the joint argument.

Remark 7.4.5. It remains unclear whether the modified cross-gradients functional can be effectively used to align gradients. Note the similarity of the modified cross-gradients with the singular values of the Jacobian in section 5.5. ▲

Proof of lemma 7.4.4. The partial derivatives of the modified cross-gradients functional are given by

$$\partial_{x_n} f(x, y) = (2|x|^2 + |y|^2)x_n - \langle x, y\rangle y_n \quad (7.33)$$

$$\partial_{x_m, x_n} f(x, y) = (2|x|^2 + |y|^2)\delta_{n,m} + 4x_n x_m - y_n y_m \quad (7.34)$$

$$\partial_{y_m, x_n} f(x, y) = 2x_n y_m - y_n x_m - \langle x, y\rangle \delta_{n,m} \quad (7.35)$$

so that the Hessian $Hf(x, y)$ in the joint argument reads

$$\begin{pmatrix} (2|x|^2 + |y|^2)I + 4xx^T - yy^T & 2xy^T - yx^T - \langle x, y\rangle I \\ 2xy^T - yx^T - \langle x, y\rangle I & (|x|^2 + 2|y|^2)I + 4yy^T - xx^T \end{pmatrix}. \quad (7.36)$$

We see that the Hessian $Hf(x,y)$ is positive-semidefinite. Indeed,

$$[z_1;z_2]^T Hf(x,y)[z_1;z_2] \tag{7.37}$$

$$\begin{aligned}
&= 2\left|x\right|^2\left|z_1\right|^2 + \left|y\right|^2\left|z_1\right|^2 + 4\left\langle x,z_1\right\rangle^2 - \left\langle y,z_1\right\rangle^2 \\
&\quad + 2\left\langle x,z_1\right\rangle\left\langle y,z_2\right\rangle - \left\langle x,z_2\right\rangle\left\langle y,z_1\right\rangle - \left\langle x,y\right\rangle\left\langle z_1,z_2\right\rangle \\
&\quad + 2\left\langle x,z_2\right\rangle\left\langle y,z_1\right\rangle - \left\langle x,z_1\right\rangle\left\langle y,z_2\right\rangle - \left\langle x,y\right\rangle\left\langle z_1,z_2\right\rangle \\
&\quad + 2\left|y\right|^2\left|z_2\right|^2 + \left|x\right|^2\left|z_2\right|^2 + 4\left\langle y,z_2\right\rangle^2 - \left\langle x,z_2\right\rangle^2
\end{aligned} \tag{7.38}$$

$$\begin{aligned}
&= \underbrace{\left|y\right|^2\left|z_1\right|^2 - \left\langle y,z_1\right\rangle^2}_{\geq 0} + \underbrace{\left|x\right|^2\left|z_2\right|^2 - \left\langle x,z_2\right\rangle^2}_{\geq 0} + \underbrace{4\left\langle x,z_1\right\rangle^2 + 4\left\langle y,z_2\right\rangle^2}_{\geq 0} \\
&\quad + 2\left|x\right|^2\left|z_1\right|^2 + 2\left|y\right|^2\left|z_2\right|^2 \\
&\quad + \underbrace{\left\langle x,z_1\right\rangle\left\langle y,z_2\right\rangle - 2\left\langle x,y\right\rangle\left\langle z_1,z_2\right\rangle + \left\langle x,z_2\right\rangle\left\langle y,z_1\right\rangle}_{\geq -4|x||z_1||y||z_2|}
\end{aligned} \tag{7.39}$$

$$\geq 2\left|x\right|^2\left|z_1\right|^2 + 2\left|y\right|^2\left|z_2\right|^2 - 4\left|x\right|\left|z_1\right|\left|y\right|\left|z_2\right| = 2(\left|x\right|\left|z_1\right| - \left|y\right|\left|z_2\right|)^2 \geq 0. \tag{7.40}$$

■

Example 7.4.6 (Quadratic Parallel Level Sets). The quadratic parallel level sets functional is convex in each but not convex in the joint argument. ▲

Proof. The local regularization function for quadratic parallel level sets is

$$f_{\mathrm{PL}_\beta^{\mathcal{Q}}}(x,y) = \left[\beta^4 + \beta^2(\left|x\right|^2 + \left|y\right|^2) + \left|x\right|^2\left|y\right|^2 - \left\langle x,y\right\rangle^2\right]^{1/2}. \tag{7.41}$$

To proof the first claim, observe that with $A = (\left|y\right|^2 + \beta^2)I - yy^T$, similarly to lemma 5.4.3 there is

$$f_{\mathrm{PL}_\beta^{\mathcal{Q}}}(x,y) = (\beta^4 + \beta^2\left|y\right|^2 + \left|Ax\right|^2)^{1/2}. \tag{7.42}$$

As $f(s) = (c+s^2)^{1/2}, c \geq 0$ is convex and monotonically increasing, the convexity in x follows from lemma 7.1.5.

For the non-convexity in the joint argument, assume that $f_{\mathrm{PL}_\beta^{\mathcal{Q}}}$ is convex. Then it follows from theorem 7.3.1 that for all x and y there is

$$\begin{aligned}
&2\Big[\beta^4 + \beta^2(\left|\tfrac{1}{2}x\right|^2 + \left|\tfrac{1}{2}y\right|^2) + \left|\tfrac{1}{2}x\right|^2\left|\tfrac{1}{2}y\right|^2 - \left\langle\tfrac{1}{2}x,\tfrac{1}{2}y\right\rangle^2\Big]^{1/2} \\
&\qquad\qquad \leq (\beta^4 + \beta^2\left|x\right|^2)^{1/2} + (\beta^4 + \beta^2\left|y\right|^2)^{1/2}.
\end{aligned} \tag{7.43}$$

In particular for $x = sx_0, y = sy_0, \langle x_0, y_0\rangle = 0, |x_0|, |y_0| = 1$ and $s > \sqrt{8}\beta$ there is

$$2\big(\beta^4 + 2^{-1}\beta^2 s^2 + 2^{-4}s^4\big)^{1/2} \leq 2\big(\beta^4 + \beta^2 s^2\big)^{1/2} \tag{7.44}$$

and hence $\frac{1}{8}s^2 \leq \beta^2$ which contradicts $s > \sqrt{8}\beta$. ■

Example 7.4.7 (Linear Parallel Level Sets). Linear parallel level sets is not convex in the joint argument. It is also not convex in each argument provided the other argument is non-zero. ▲

Proof. The local regularization function for linear parallel level sets is

$$f_{\mathrm{PL}^{\mathcal{L}}_{\beta}}(x, y) = \big(|x|^2 + \beta^2\big)^{1/2}\big(|y|^2 + \beta^2\big)^{1/2} - \big(\langle x, y\rangle^2 + \beta^4\big)^{1/2}. \tag{7.45}$$

Denoting $f(t) = \big[(1+\beta^2)t^2 + 1 + 2\beta^2 + \beta^4\big]^{1/2} - \big[(1+\beta^2)t^2 + \beta^2 + \beta^4\big]^{1/2}$ there is $\lim_{t\to\infty} f(t) = 0$. Indeed,

$$\begin{aligned} 0 \leq f(t) &= \big[(1+\beta^2)t^2 + 1 + 2\beta^2 + \beta^4\big]^{1/2} - \big[(1+\beta^2)t^2 + \beta^2 + \beta^4\big]^{1/2} && (7.46)\\ &= \frac{(1+\beta^2)t^2 + 1 + 2\beta^2 + \beta^4 - \big[(1+\beta^2)t^2 + \beta^2 + \beta^4\big]}{\big[(1+\beta^2)t^2 + 1 + 2\beta^2 + \beta^4\big]^{1/2} + \big[(1+\beta^2)t^2 + \beta^2 + \beta^4\big]^{1/2}} && (7.47)\\ &< \frac{(1+\beta^2)^{1/2}}{t} \to 0 && (7.48) \end{aligned}$$

Therefore, there exists t_0 such that $f(t_0) < (\beta^4 + 1)^{1/2} - \beta^2$. Motivated by figure 5.3, let $x(s) \coloneqq x_0 + sy/|y|$ with $\langle x_0, y\rangle = 0, |x_0| = t_0$ and

$$\begin{aligned} g(s) \coloneqq f_{\mathrm{PL}^{\mathcal{L}}_{\beta}}(x(s), y/|y|) &= \big(t_0^2 + s^2 + \beta^2\big)^{1/2}\big(1+\beta^2\big)^{1/2} - \big(s^2 + \beta^4\big)^{1/2} && (7.49)\\ &= \big[(1+\beta^2)t_0^2 + s^2 + \beta^2 s^2 + \beta^2 + \beta^4\big]^{1/2} - \big(s^2 + \beta^4\big)^{1/2} && (7.50) \end{aligned}$$

and hence

$$\begin{aligned} g(\pm 1) - g(0) =& \big[(1+\beta^2)t_0^2 + 1 + 2\beta^2 + \beta^4\big]^{1/2}\\ &- \big[(1+\beta^2)t_0^2 + \beta^2 + \beta^4\big]^{1/2} - \big[(1+\beta^4)^{1/2} - \beta^2\big] && (7.51)\\ =& f(t_0) - \big[(1+\beta^4)^{1/2} - \beta^2\big] < 0. && (7.52) \end{aligned}$$

Thus, g is not convex and it follows that $f_{\mathrm{PL}^{\mathcal{L}}_{\beta}}$ is not convex in each argument. As it is not convex in each argument it cannot be convex in the joint argument. Similarly to example 7.4.6, we could prove this also directly. ■

Table 7.1. *Summary of convexity results.*

method	convex in each argument	convex in joint argument	promotes joint structure
$\mathbb{TV}_\beta^{\mathcal{S}}$	yes	yes	no
$\mathbb{TV}_\beta^{\mathcal{J}}$	yes	yes	yes
CG	yes	no	yes
modified CG	yes	yes	?
$\mathrm{PL}_\beta^{\mathcal{Q}}$	yes	no	yes
$\mathrm{PL}_\beta^{\mathcal{L}}$	no	no	yes

Remark 7.4.8. Non-convex functionals are sometimes approximated by their bi-conjugate, i.e. (neglecting smoothness) the largest convex functional smaller than the original functional. The bi-conjugate of $f_{\mathrm{APL}^{\mathcal{L}}}$ is zero, so there is no gain by approximating $f_{\mathrm{APL}^{\mathcal{L}}}$ with its bi-conjugate. ▲

Proof. Assume there is a convex function $f\colon \mathbb{R}^D \to \mathbb{R}$ that is larger than zero, i.e. $f \geq 0$ and $f \neq 0$, and smaller than $f_{\mathrm{APL}^{\mathcal{L}}}$, i.e. $f \leq f_{\mathrm{APL}^{\mathcal{L}}}$. As f is larger than zero, there exists a point x_0 such that $f(x_0) > 0$. Let us now consider $x(s) \coloneqq x_0 + s\theta$ and notice that $f_{\mathrm{APL}^{\mathcal{L}}}(x(s)) \to 0$ as $s \to \pm\infty$. Indeed,

$$0 \leq f_{\mathrm{APL}^{\mathcal{L}}}(x(s)) = \left(|x_0 + s\theta|^2 + \beta^2\right)^{1/2} - \left(\langle x_0 + s\theta, \theta\rangle^2 + \beta^2\right)^{1/2} \tag{7.53}$$

$$= \frac{\left(|x_0|^2 + 2s\langle x_0, \theta\rangle + s^2 + \beta^2\right) - \left(\langle x_0, \theta\rangle^2 + 2s\langle x_0, \theta\rangle + s^2 + \beta^2\right)}{\left(|x_0 + s\theta|^2 + \beta^2\right)^{1/2} + \left(\langle x_0 + s\theta, \theta\rangle^2 + \beta^2\right)^{1/2}} \tag{7.54}$$

$$\leq \frac{|x_0|^2 - \langle x_0, \theta\rangle^2}{\left(|x_0 + s\theta|^2 + \beta^2\right)^{1/2}} \overset{s\to\pm\infty}{\to} 0. \tag{7.55}$$

As $0 \leq f \leq f_{\mathrm{APL}^{\mathcal{L}}}$, there is also $f(x(s)) \to 0$ as $s \to \pm\infty$. Therefore, there are $s_1 < 0 < s_2$ such that for $x_n \coloneqq x(s_n), n = 1, 2$ there is $f(x_n) < f(x_0)$. In addition, let $t \coloneqq s_2/(s_2 - s_1) > 0$. Then there is $(1 - t) = -s_1/(s_2 - s_1)$ and we have

$$x_0 = tx_0 + (1-t)x_0 + \tfrac{s_1 s_2}{s_2 - s_1}\theta + \tfrac{-s_1 s_2}{s_2 - s_1}\theta = tx_1 + (1-t)x_2. \tag{7.56}$$

With the convexity of f, there is

$$f(x_0) = f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2) \tag{7.57}$$

$$< tf(x_0) + (1-t)f(x_0) = f(x_0). \tag{7.58}$$

■

Despite the non-convexity we can show that the linear parallel level sets functional

is convex for small arguments.

Lemma 7.4.9. *Let $\beta > 0, y \neq 0, r = \beta/|y|$ with $\beta^4(r^2 + r^4) \geq 1$. In addition, let*

$$R(\beta, |y|) = \beta\{[\beta^4(r^2 + r^4)]^{1/3} - 1\}^{1/2}. \tag{7.59}$$

Then the linear parallel level sets functional is convex in x on $\mathbb{B}_{R(\beta,|y|)}$, i.e. a ball with radius $R(\beta, |y|)$ around zero.

Proof. The gradient and Hessian in the first argument of the local regularization function for linear parallel level sets are

$$\partial_{x_n} f_{\mathrm{PL}^{\mathcal{L}}_{\beta}}(x, y) = x_n |y|_\beta |x|_\beta^{-1} - y_n \langle x, y\rangle |\langle x, y\rangle|_{\beta^2}^{-1} \tag{7.60}$$

$$\begin{aligned}\partial_{x_m,x_n} f_{\mathrm{PL}^{\mathcal{L}}_{\beta}}(x, y) &= \delta_{n,m} |y|_\beta |x|_\beta^{-1} - x_n x_m |y|_\beta |x|_\beta^{-3} \\ &\quad + y_n y_m \langle x, y\rangle^2 |\langle x, y\rangle|_{\beta^2}^{-3} - y_n y_m |\langle x, y\rangle|_{\beta^2}^{-1} \qquad (7.61)\\ &= |\langle x, y\rangle|_{\beta^2}^{-3} |x|_\beta^{-3} \Big(\delta_{n,m} |y|_\beta |x|_\beta^2 |\langle x, y\rangle|_{\beta^2}^3 \\ &\quad - x_n x_m |y|_\beta |\langle x, y\rangle|_{\beta^2}^3 - y_n y_m \beta^2 |x|_\beta^3\Big) \qquad (7.62)\end{aligned}$$

Note

$$r^2 + r^4 = \left(\frac{\beta}{|y|}\right)^2 + \left(\frac{\beta}{|y|}\right)^4 = \frac{\beta^2}{|y|^4}(|y|^2 + \beta^2) = \frac{\beta^2 |y|_\beta^2}{|y|^4}. \tag{7.63}$$

For every $x \in \mathbb{B}_{R(\beta,|y|)}$ there is

$$|x|_\beta \leq \beta^{5/3}(r^2 + r^4)^{1/6} = \beta^2 \left(\frac{|y|_\beta}{|y|^2}\right)^{1/3} \tag{7.64}$$

thus, $|y|^2 |y|_\beta^{-1} |x|_\beta^3 \leq \beta^6$. Hence,

$$\begin{aligned}&|\langle x, y\rangle|_{\beta^2}^3 |x|_\beta^3 z^T H f_{\mathrm{PL}^{\mathcal{L}}_{\beta}}(x, y) z \qquad (7.65)\\ &\quad = |z|^2 |y|_\beta |x|_\beta^2 |\langle x, y\rangle|_{\beta^2}^3 - \langle z, x\rangle^2 |y|_\beta |\langle x, y\rangle|_{\beta^2}^3 - \langle z, y\rangle^2 \beta^2 |x|_\beta^3 \qquad (7.66)\\ &\quad = \underbrace{|y|_\beta |\langle x, y\rangle|_{\beta^2}^3 (|z|^2 |x|^2 - \langle z, x\rangle^2)}_{\geq 0} + |z|^2 |y|_\beta \beta^2 |\langle x, y\rangle|_{\beta^2}^3 - \langle z, y\rangle^2 \beta^2 |x|_\beta^3 \qquad (7.67)\\ &\quad \geq |z|^2 \beta^2 \Big(|y|_\beta |\langle x, y\rangle|_{\beta^2}^3 - |y|^2 |x|_\beta^3\Big) \geq |z|^2 \beta^2 |y|_\beta \Big(\beta^6 - |y|^2 |y|_\beta^{-1} |x|_\beta^3\Big) \geq 0. \qquad (7.68)\end{aligned}$$

■

Remark 7.4.10. Let $|y| = \sqrt{\sqrt{33} + 1}/4 \approx 0.65$ and $\beta = 1$. The pair fulfils the

conditions of the lemma and there is convexity on the unit ball, i.e. $|x| \leq 1$. ▲

7.5. Summary.

In this chapter we have analysed the priors for joint reconstruction in terms of convexity. It turns out that the considered examples of parallel level set priors are all not convex in the joint argument as they fail the growth condition. For quadratic parallel level sets, there is at least convexity in each argument.

Chapter 8

Applications in Colour Imaging

In this chapter we will present the application of the joint reconstruction framework to RGB colour imaging. We have seen in figure 1.2 that the colour channels of RGB images exhibit a strong correlation. We will exploit the common structure of the colour channels for denoising and demosaicking. This chapter has been published in [50].

8.1. Problem Set-up.

The two tasks in RGB colour imaging we will consider here are denoising and demosaicking. In the case of *denoising* the problem is restoration of a noisy input colour image. It is often assumed that we have knowledge about the type of noise. In our case, we will perform denoising of colour images that have been corrupted with Gaussian noise.

The problem of demosaicking arises naturally after image acquisition. The colour channels of an RGB image are usually acquired separately, but as each detector element in a camera can only acquire the data for one colour channel, at each spatial position there is only data for one colour available. A typical arrangement of the detectors for the three colour channels in an RGB camera is the so called *Bayer filter* shown in figure 8.1 on the left and a sample image acquired with this filter on the right. As we want to form an image out of these partial measurements, we will formulate this as an inverse

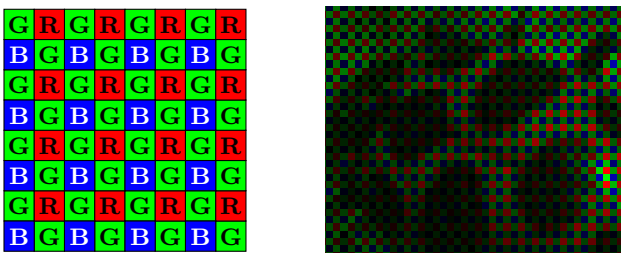


Figure 8.1. *Left: The Bayer filter for demosaicking. At each pixel only one detector element acquires either information for the red, green or blue channel so that reconstruction is needed to get the full image. Right: Acquired data with the Bayer filter. Note that non-acquired data is set to zero in order to visualize it as an RGB image.*

problem and reconstruct an image that matches the data and fits our a priori knowledge about images from natural scenes.

8.2. Material and Methods.

Methods. We will compare colour total variation, cf. example 4.3.9, with quadratic and linear parallel level sets, cf. examples 5.2.3 and 5.2.2. The two parallel level set priors are extended to three dimensions according to remark 5.2.8. In every case we want to minimize

$$\mathcal{J}(w) = \sum_{n\in\{r,g,b\}} \left\{ |C_n w_n - f_n|^2 \right\} + \alpha \mathcal{P}(w), \tag{8.1}$$

where $w = (w_r, w_g, w_b)$ is an RGB colour image and the prior $\mathcal{P}(w)$ is specified according to the method. For denoising, the operators are the identity, i.e. $C_n = I, n \in \{r, g, b\}$, and in the case of demosaicking they perform a projection that matches the Bayer filter.

Numerical Implementation. All three functionals are implemented in the same manner and in every case the functional is minimized by a Quasi-Newton method. To be more precise, we used a large scale version of BFGS [117, p. 226]. For a Quasi-Newton method, we need to provide a black box function that computes the function value and its gradient for any arbitrary input image. Therefore, we need to discretize the prior $\mathcal{P}$ and its gradient / Gâteaux derivative. We will use the forward-backward discretization we have presented in section 6.3. The MATLAB implementation can be found on [49].

Data. We took five test images for this experiment from the Berkeley Segmentation database [107, 2] and used the noisy versions from [58]. The noise was modelled to be additive, uncorrelated Gaussian noise of standard deviations 5, 10, 15, 25 and 35 for images in the colour range of [0, 255].

Parameter Choice. We varied the overall regularization parameter α as well as the smoothing parameter β for all three methods. As this is a simulation, we know the ground truth in every case and we can therefore choose "optimal" parameters. The parameters are chosen to maximize the peak signal-to-noise ratio (PSNR) between our

estimated image $w^\sharp$ and the ground truth image w_0, which we define as

$$\mathrm{PSNR}(w^\sharp, w_0) := 10 \log_{10} \left(\frac{255^2}{|w^\sharp - w_0|^2} \right). \tag{8.2}$$

This is a logarithmic measure and as such it is usually referred to in units of dB.

Another "parameter" is the initial guess, as we are minimizing non-convex functionals with gradient-based optimization methods. We set the initial image to be the input data, i.e. in the case of denoising the initial guess is chosen to be the noisy input image and for demosaicking the initial guess is the zero-filled data as shown in figure 8.1. The results show that this is a sufficiently good choice for these kind of problems.

Evaluation. We will evaluate the results i) visually and ii) quantitatively in terms of their PSNR and SSIM. The SSIM (Structural SIMilarity index) [156] also measures the similarity of an image $w^\sharp$ to another image w_0 but is known to be closer to human perception than the PSNR. It is bounded from above by one for perfectly matching images and as such we will refer to it in percentages. We will use the SSIM function for MATLAB that has been made available online [157].

8.3. Results for Denoising of Colour Images.

The visual results for the test image `bugs` are shown in figures 8.2 - 8.4. We can see that all methods perform well for a small noise level. When the noise level increases, colour total variation shows artefactual colour fluctuations. While colour total variation and quadratic parallel level sets smooth the images a lot to cope with the noise, linear parallel level sets yields sharper images. An interesting observation is that the red colour of the bug seems to be better preserved by colour total variation than the two other methods.

The quantitative results in figure 8.5 are similar to our previous observations. For all of five cases of noise, the mean PSNR and mean SSIM taken over the five test images show that the two parallel level set based priors clearly outperform colour total variation.

Finally, we plot the mean choice of the regularization parameter α and the smoothing parameter β in figure 8.6. As expected, the regularization parameter α is increasing with increasing amount of noise for all three methods. While for colour total variation and quadratic parallel level sets the parameter β is almost constant, it increases with noise for linear parallel level sets.

Figure 8.2. *Denoising of test image* `bugs`*. Comparing the results for colour total variation ($\mathbb{TV}^{\mathcal{C}}_{\beta}$), quadratic parallel level sets ($\mathrm{PL}^{\mathcal{Q}}_{\beta}$) and linear parallel level sets ($\mathrm{PL}^{\mathcal{L}}_{\beta}$) for an increasing level of noise (std). The parameters are chosen to maximize the peak signal-to-noise ratio.*

Figure 8.3. *Close-up of figure 8.2 showing denoising results for test image* `bugs`.

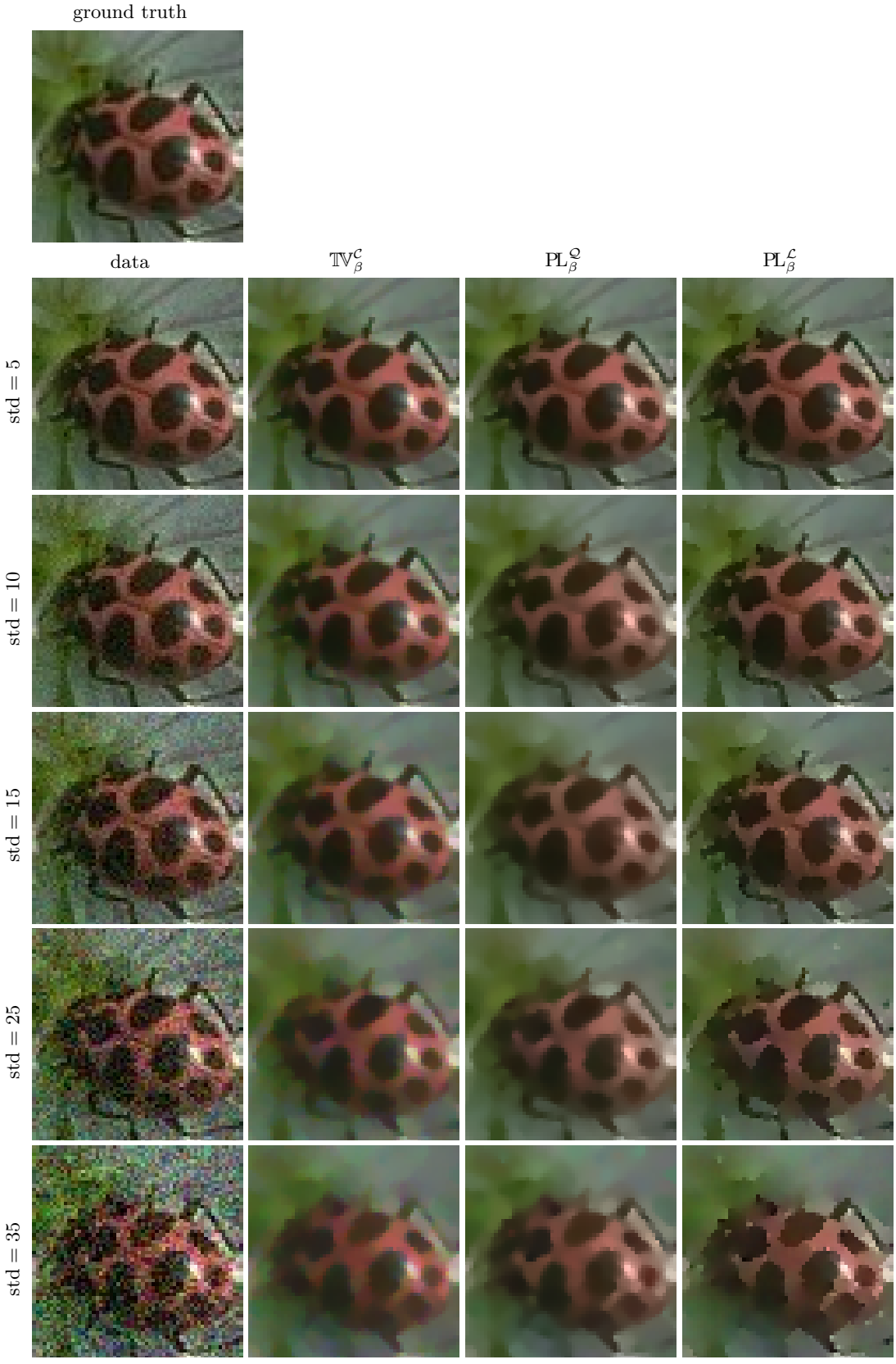


Figure 8.4. *Close-up of figure 8.2 showing denoising results for test image* `bugs`.

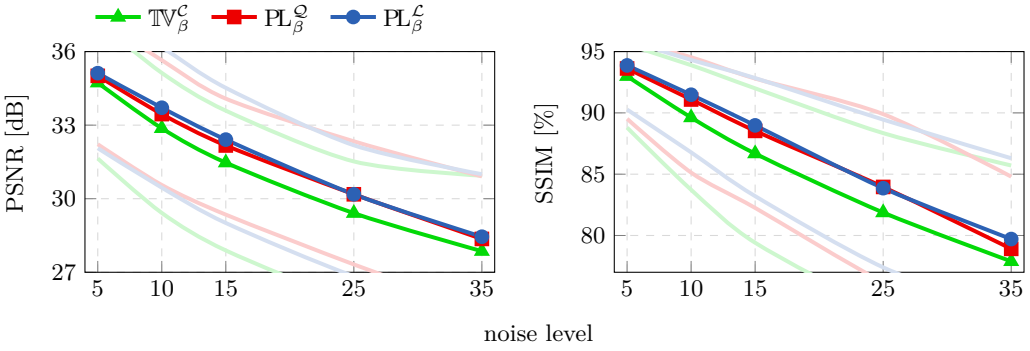


Figure 8.5. *Quantitative denoising results in terms of peak signal-to-noise (PSNR) and structural similarity index (SSIM). The solid line indicates the mean and the dotted lines the minimal and maximal PSNR and SSIM taken over all five test images.*

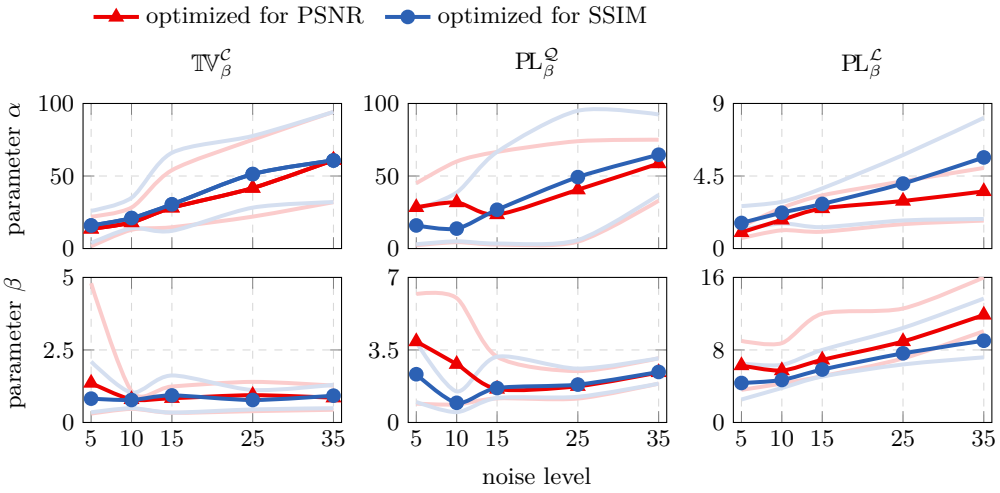


Figure 8.6. *Parameters for denoising results optimized to maximize the peak signal-to-noise ratio (PSNR; red) and the structural similarity index (SSIM; blue). Plotted are the mean values and shaded the maximum and minimum values of the parameters.*

8.4. Results for Demosaicking of Colour Images.

Visual results for demosaicking of colour images are presented in figures 8.7 - 8.9. Looking at the whole images as such in figure 8.7, all methods give reasonable results. When we look at a close-up of these images on fine details, such as the tree in figure 8.9, we can see a big difference between the methods. While colour total variation gives distracting colour artefacts, both parallel level set methods give reasonable reconstructions. For increasing noise level, hence increasing regularization parameter, the difference between quadratic and linear parallel level sets becomes clear: the quadratic version results in smooth images and the linear one in sharp and patchy images.

More examples of demosaicking of colour images with linear parallel level sets are shown in figures 8.10 - 8.13. They are all reconstructed without visually disturbing colour artefacts.

As in the case of denoising, the quantitative results underline the visual impression, cf. figure 8.14. From figure 8.15 we can see that the mean parameters for these experiments follow the same trend: they are all increasing with respect to the noise.

8.5. Summary.

We have presented several results for denoising and demosaicking of colour images. The results for the two parallel level set based priors were in both cases superior to the results of colour total variation, both visually and quantitatively. In addition, we observed consistently that linear parallel level sets yielded sharper and better defined images than the other two methods.

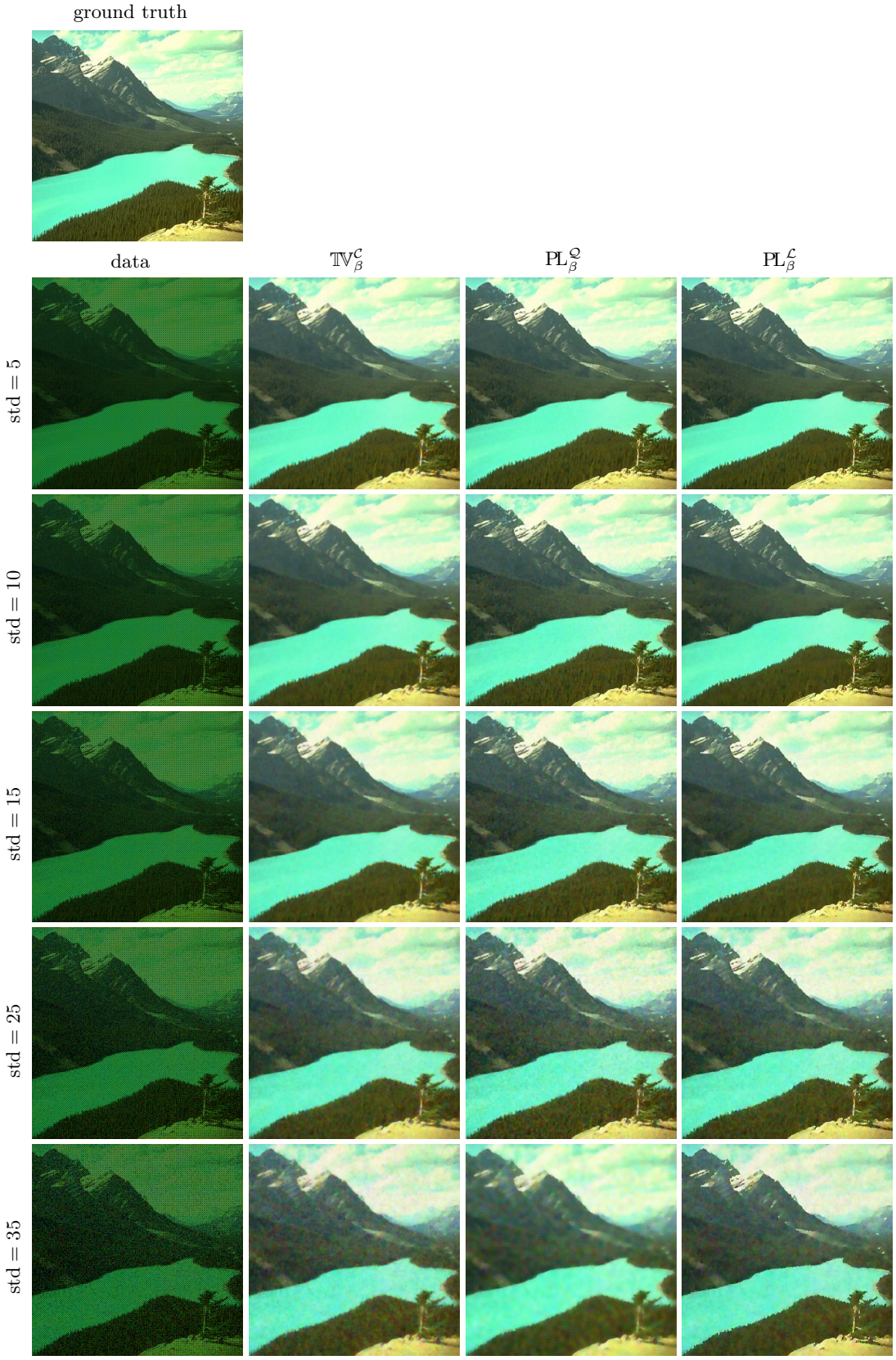


Figure 8.7. *Demosaicking of test image* `lake`*. Comparing the results for colour total variation ($\mathbb{TV}^{\mathcal{C}}_{\beta}$), quadratic parallel level sets ($\mathrm{PL}^{\mathcal{Q}}_{\beta}$) and linear parallel level sets ($\mathrm{PL}^{\mathcal{L}}_{\beta}$) for an increasing level of noise (std). The parameters are chosen to maximize the peak signal-to-noise ratio.*

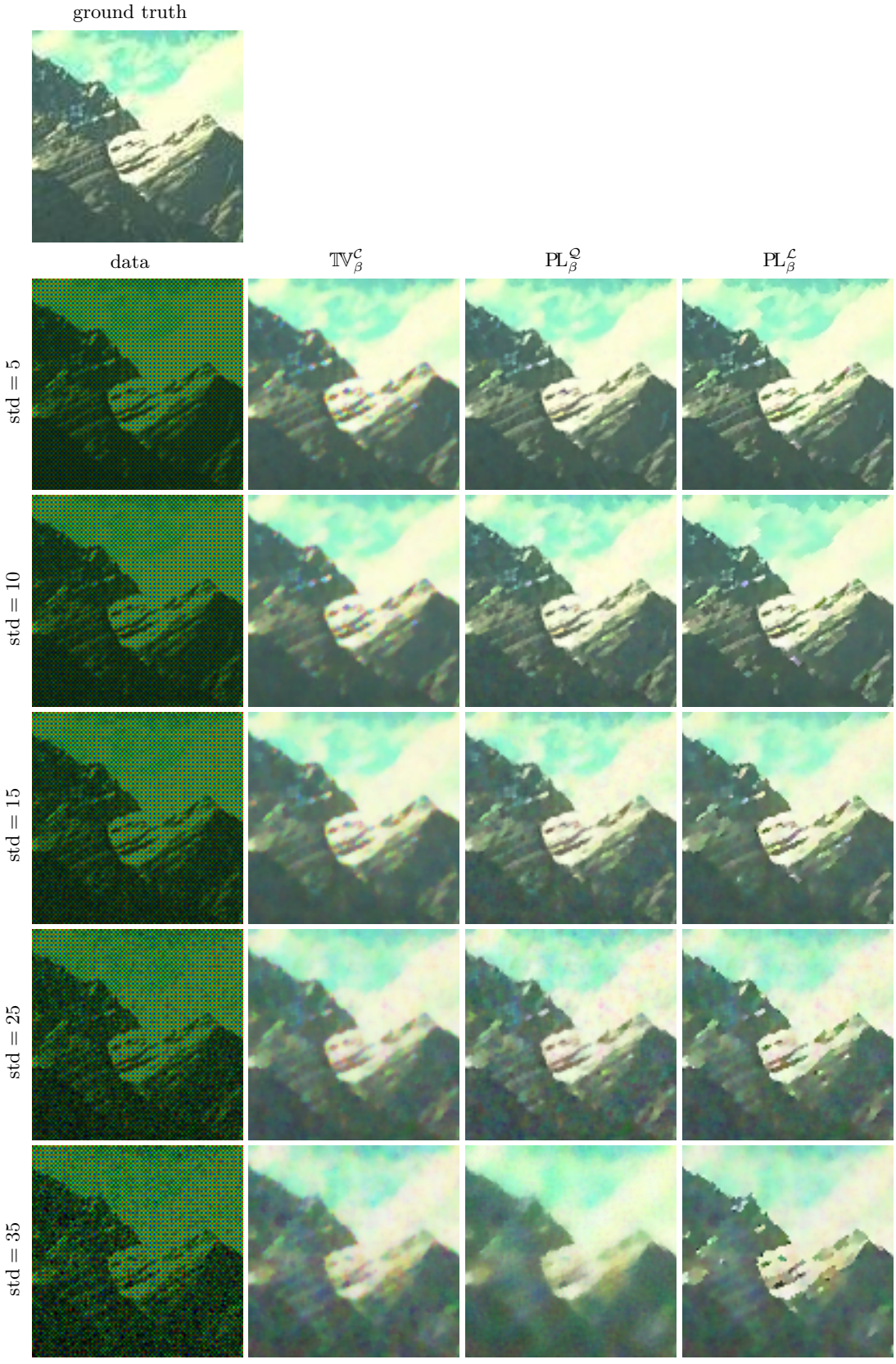


Figure 8.8. *Close-up of figure 8.7 showing demosaicking results for test image* `lake`.

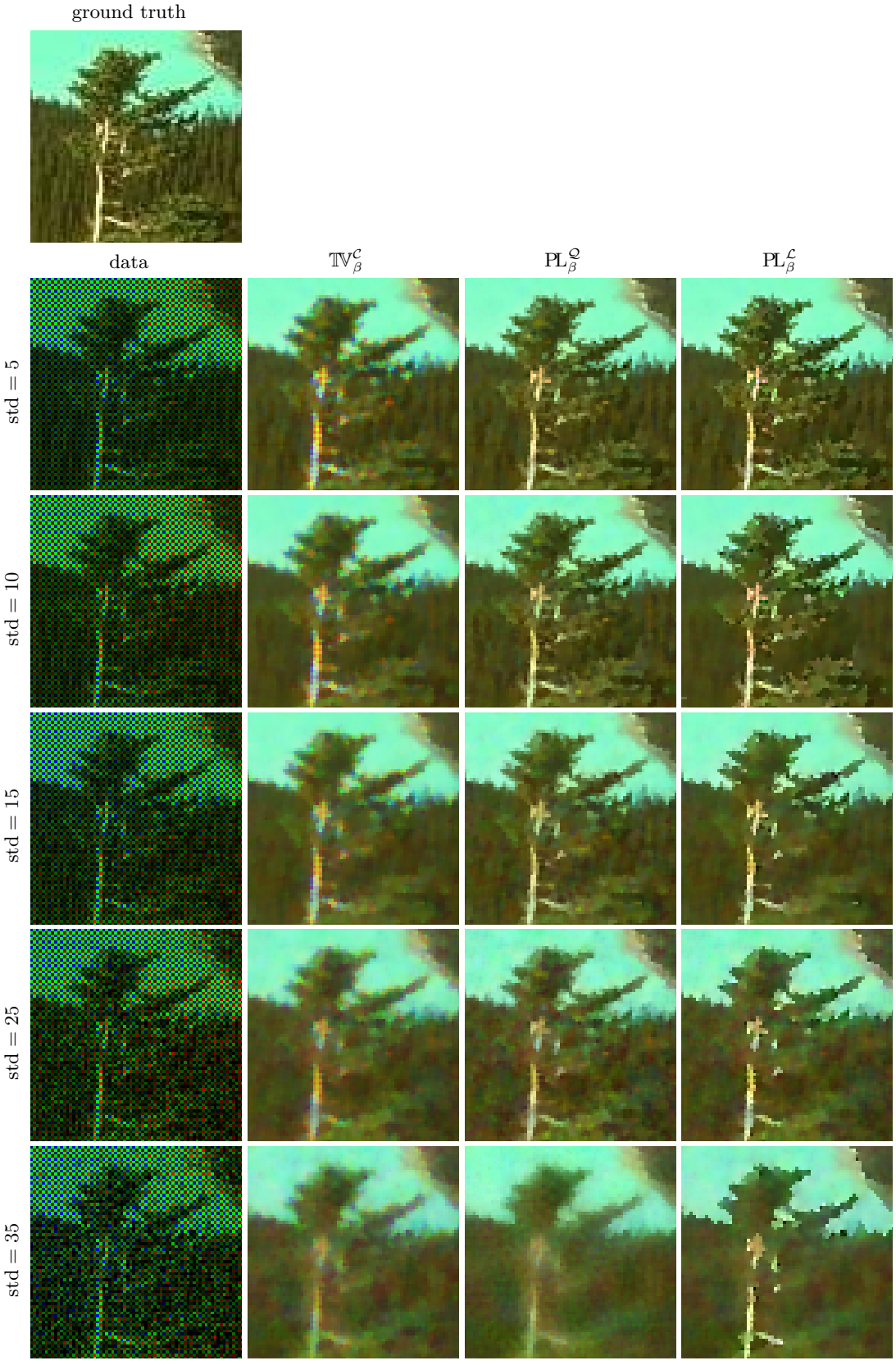


Figure 8.9. *Close-up of figure 8.7 showing demosaicking results for test image* `lake`.

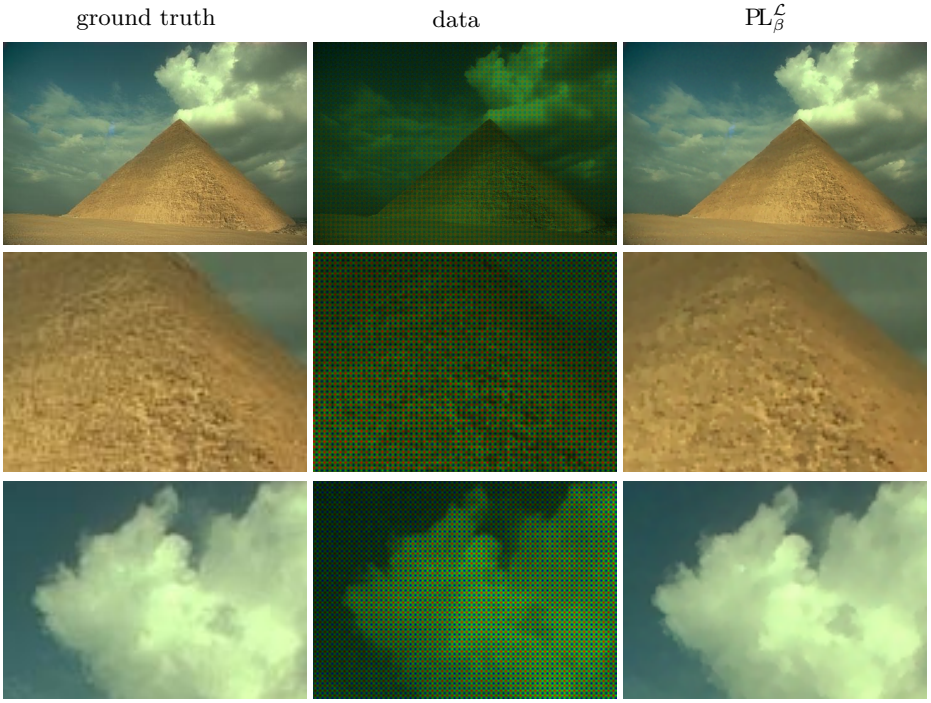


Figure 8.10. *Demosaicking results for linear parallel level sets* ($\mathrm{PL}_\beta^{\mathcal{L}}$) *with test image* `pyramid`. *The parameters are chosen to maximize the peak signal-to-noise ratio.*

Figure 8.11. *Demosaicking results for linear parallel level sets* ($\mathrm{PL}_\beta^{\mathcal{L}}$) *with test image* `wolf`. *The parameters are chosen to maximize the peak signal-to-noise ratio.*

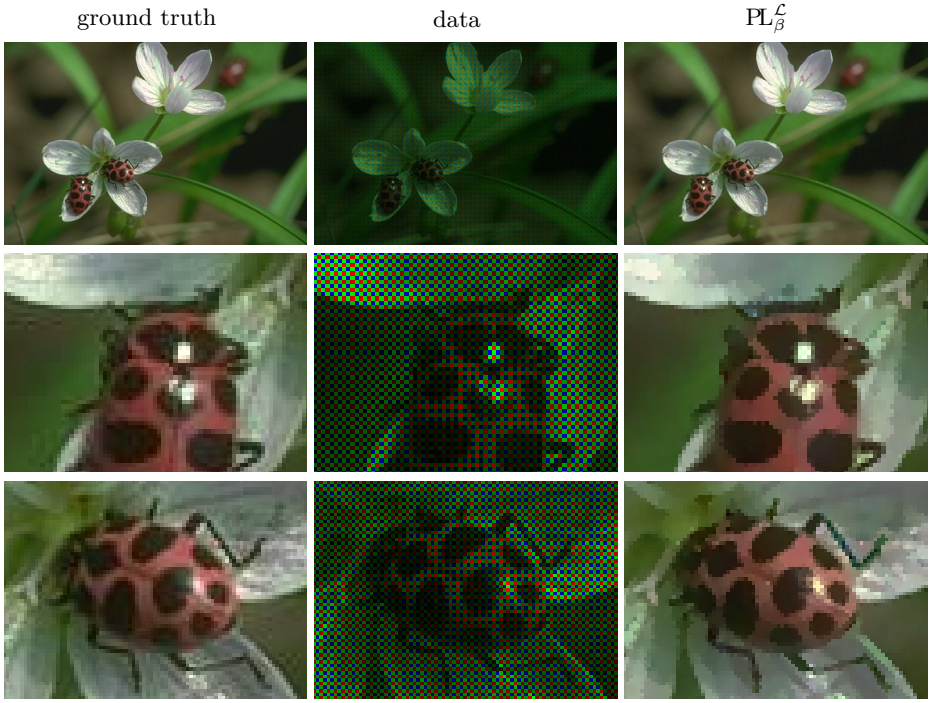


Figure 8.12. *Demosaicking results for linear parallel level sets ($\mathrm{PL}_\beta^{\mathcal{L}}$) with test image* `bugs`*. The parameters are chosen to maximize the peak signal-to-noise ratio.*

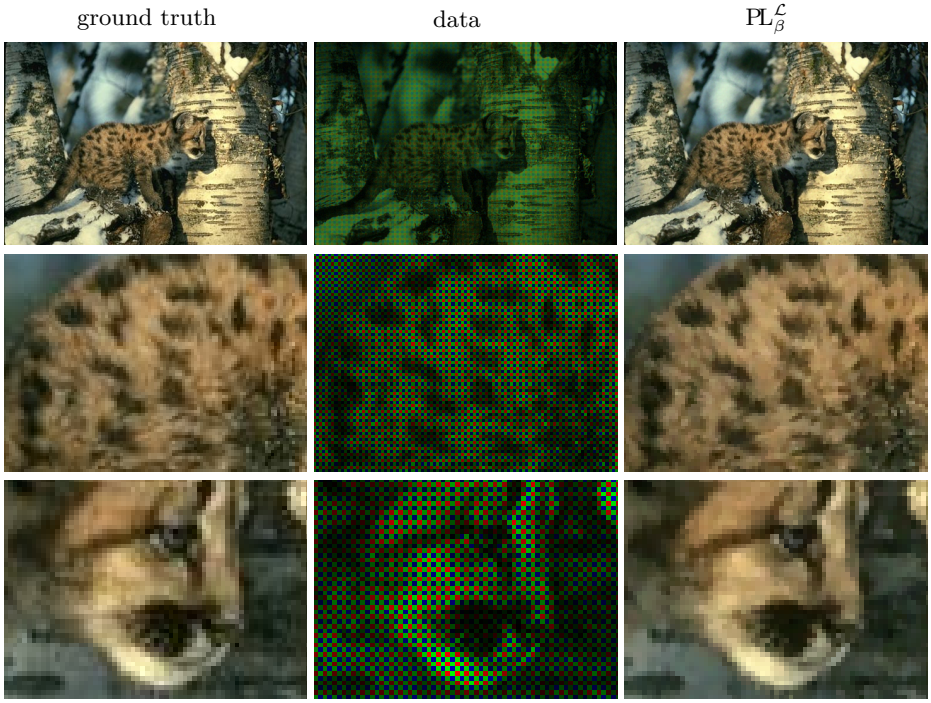


Figure 8.13. *Demosaicking results for linear parallel level sets ($\mathrm{PL}_\beta^{\mathcal{L}}$) with test image* `leopard`*. The parameters are chosen to maximize the peak signal-to-noise ratio.*

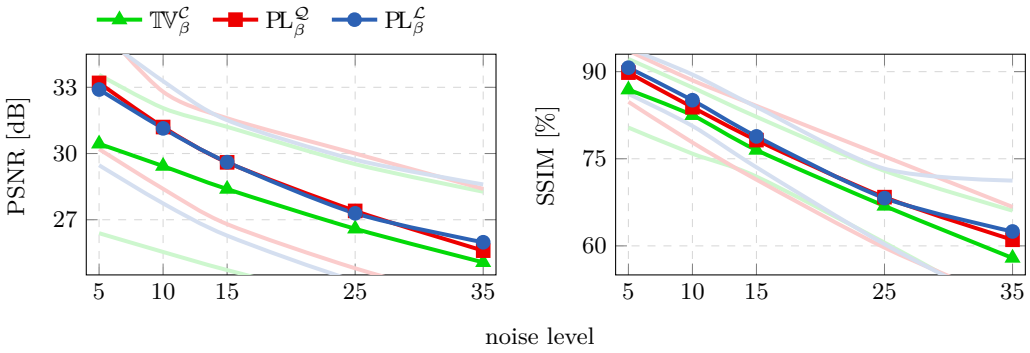


Figure 8.14. *Quantitative demosaicking results in terms of peak signal-to-noise (PSNR) and structural similarity index (SSIM). The solid line indicates the mean and the dotted lines the minimal and maximal PSNR and SSIM taken over the five test images.*

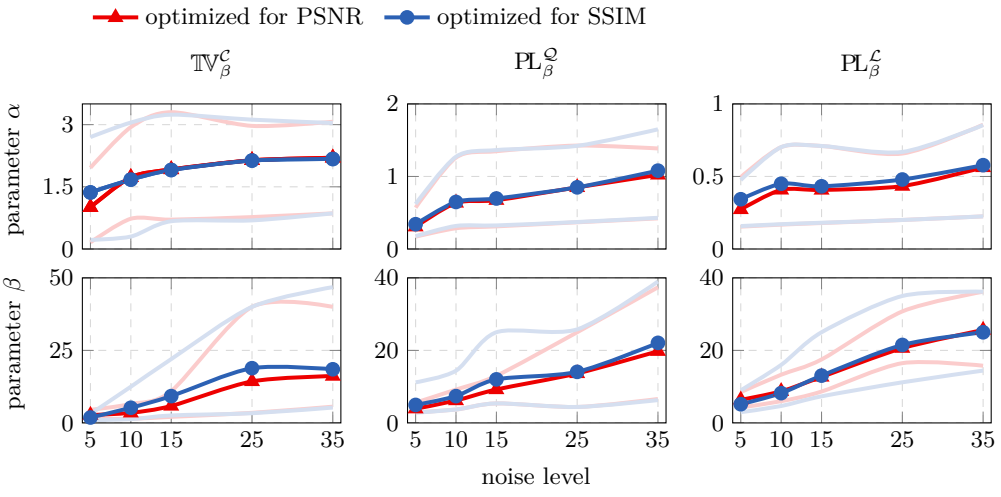


Figure 8.15. *Parameters for demosaicking results optimized to maximize the peak signal-to-noise ratio (PSNR; blue) and the structural similarity index (SSIM; red). Plotted are the mean values and shaded the maximum and minimum values of the parameters.*

Chapter 9

Applications in Medical Imaging

This chapter is dedicated to the application of joint reconstruction to medical imaging. We will first show results of joint reconstruction for PET-MRI based on results from [52]. In a realistic application, MRI images are often modelled as complex images with magnitude and phase. We will discuss the implications to MRI-only reconstruction first and then present results for this model to joint reconstruction of PET-MRI. The latter has been published in [51].

9.1. Joint Reconstruction for PET-MRI.

In this section we will present the numerical results on joint reconstruction for PET-MRI partly published in [52].

9.1.1. Material and Methods.

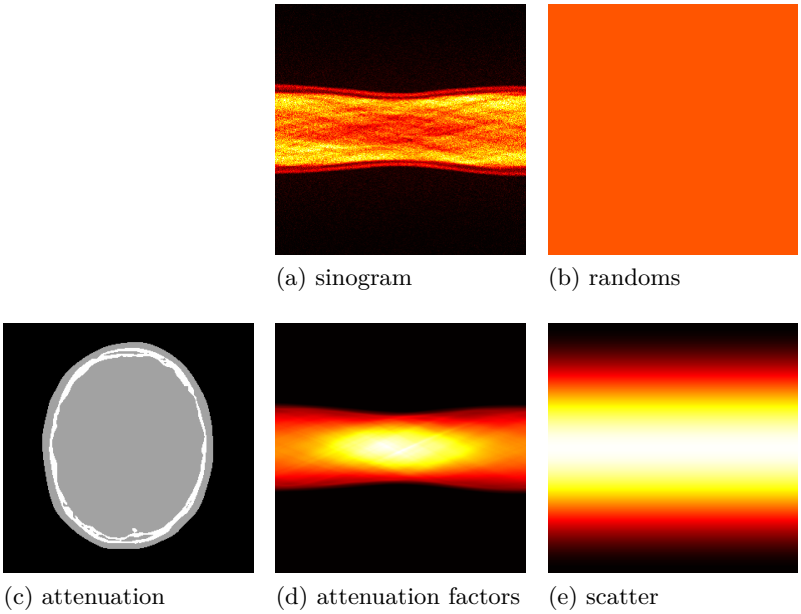
Methods. In this example we will compare separate regularization techniques: early stopping and total variation, cf. example 4.1.2, to joint regularization techniques: joint total variation, cf. example 4.1.3, quadratic parallel level sets, cf. example 5.2.3, and linear parallel level sets, cf. example 5.2.2. In addition, we present preliminary results for quadratic asymmetric parallel level sets, cf. example 5.2.10.

We perform separate PET and MRI reconstruction with total variation by minimizing

$$\mathcal{J}_{\mathrm{PET}}(u) = \mathcal{D}(Au + b|f) + \alpha \mathbb{TV}_{\beta}(u) \tag{9.1}$$

$$\mathcal{J}_{\mathrm{MRI}}(v) = \mathcal{E}(Bv|g) + \alpha \mathbb{TV}_{\beta}(v). \tag{9.2}$$

For early stopping, we set the regularization α to zero and we regularize the inversion by an early termination of the minimization algorithm.



(a) sinogram (b) randoms

(c) attenuation (d) attenuation factors (e) scatter

Figure 9.1. *Data for PET-MRI experiment with brain phantom. The range of the colour maps are (a) [0,100], (b) [0,2], (c) [0,0.015], (d) [1,12] and (e) [0, 4].*

In the case of one-sided reconstruction with quadratic asymmetric parallel level sets, we minimize

$$\mathcal{J}_{\text{PET}}(u) = \mathcal{D}(Au + b|f) + \alpha\, \text{APL}_\beta^{\mathcal{Q}}(u|v^\sharp) \tag{9.3}$$

$$\mathcal{J}_{\text{MRI}}(v) = \mathcal{E}(Bv|g) + \alpha\, \text{APL}_\beta^{\mathcal{Q}}(v|u^\sharp) \tag{9.4}$$

with the side-information $u^\sharp$ and $v^\sharp$.

For joint reconstruction, we aim to minimize

$$\mathcal{J}_{\text{PET-MRI}}(u, v) = \mathcal{D}(Au + b|f) + \mathcal{E}(Bv|g) + \alpha\mathcal{P}(u, v) \tag{9.5}$$

with the prior being either $\mathbb{TV}_\beta^{\mathcal{J}}$ for joint total variation, $\text{PL}_\beta^{\mathcal{Q}}$ for quadratic and $\text{PL}_\beta^{\mathcal{L}}$ for linear parallel level sets.

Remark 9.1.1. A similar approach to (9.5) with more than one prior has been used recently for joint reconstruction of PET-MRI as well [94, 92]. In their work, next to separate priors of the modalities, they exploit the expected common support of the wavelet coefficients of the images of the two modalities. It has been out of the scope of this work to directly compare these two approaches. Nevertheless, smooth joint total variation is related to this approach for the Haar wavelet as we discretize the spatial derivatives with forward differences. ▲

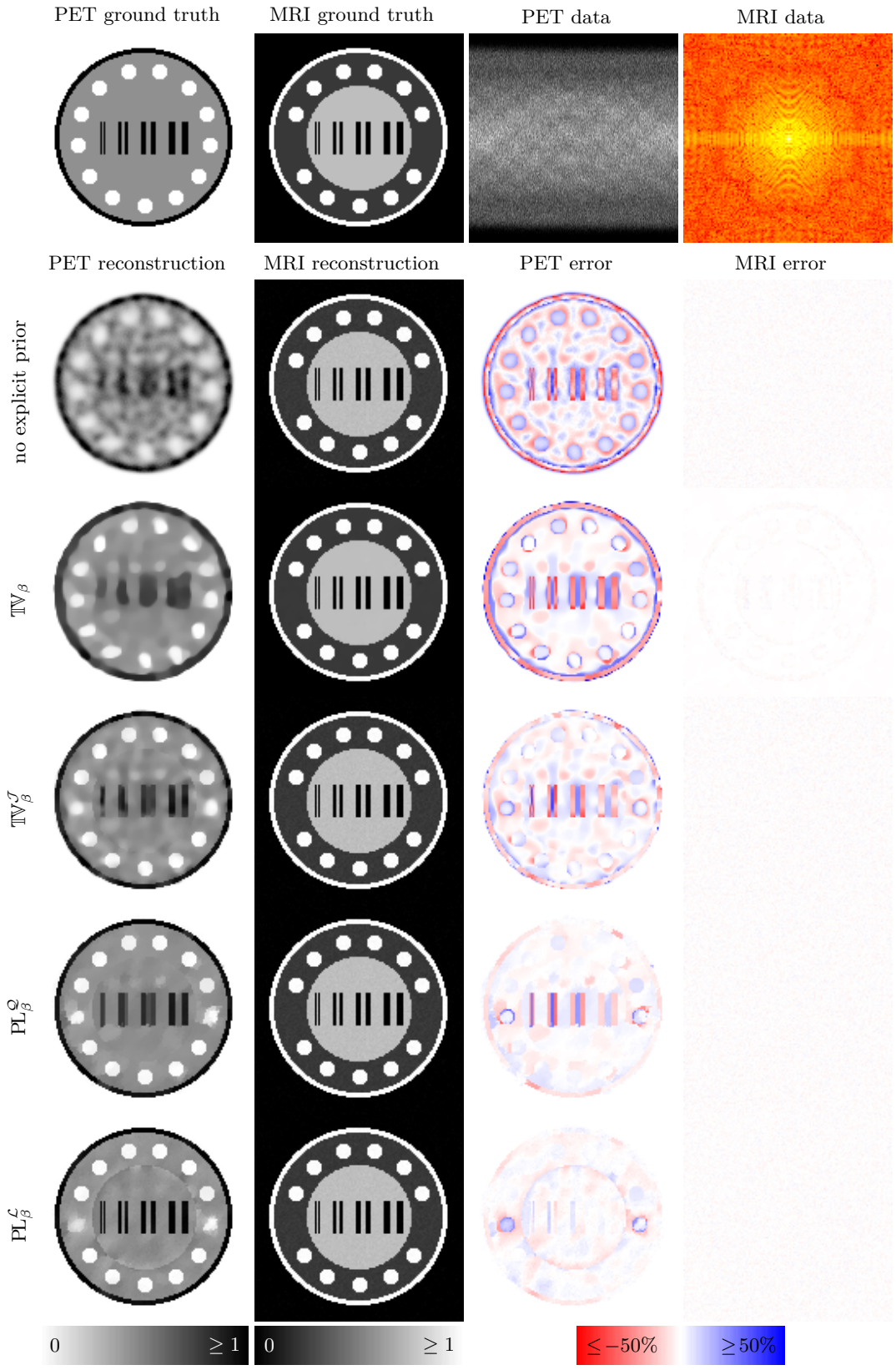


Figure 9.2. *Results for joint reconstruction of PET-MRI with the data set* `full`*, i.e. MRI is fully sampled, for early stopping / zero filling (no explicit prior), total variation ($\mathbb{TV}_\beta$), joint total variation ($\mathbb{TV}^{\mathcal{J}}_\beta$), quadratic parallel level sets ($\mathrm{PL}^{\mathcal{Q}}_\beta$) and linear parallel level sets ($\mathrm{PL}^{\mathcal{L}}_\beta$). Parameters are chosen to minimize the relative ℓ^2-error.*

Figure 9.3. *Results for joint reconstruction of PET-MRI with data set* `radial20`*, i.e. MRI is sampled along twenty radial spokes, for early stopping / zero filling (no explicit prior), total variation ($\mathbb{TV}_\beta$), joint total variation ($\mathbb{TV}_\beta^{\mathcal{J}}$), quadratic parallel level sets ($\mathrm{PL}_\beta^{\mathcal{Q}}$) and linear parallel level sets ($\mathrm{PL}_\beta^{\mathcal{L}}$). Parameters are chosen to minimize the relative ℓ^2-error.*

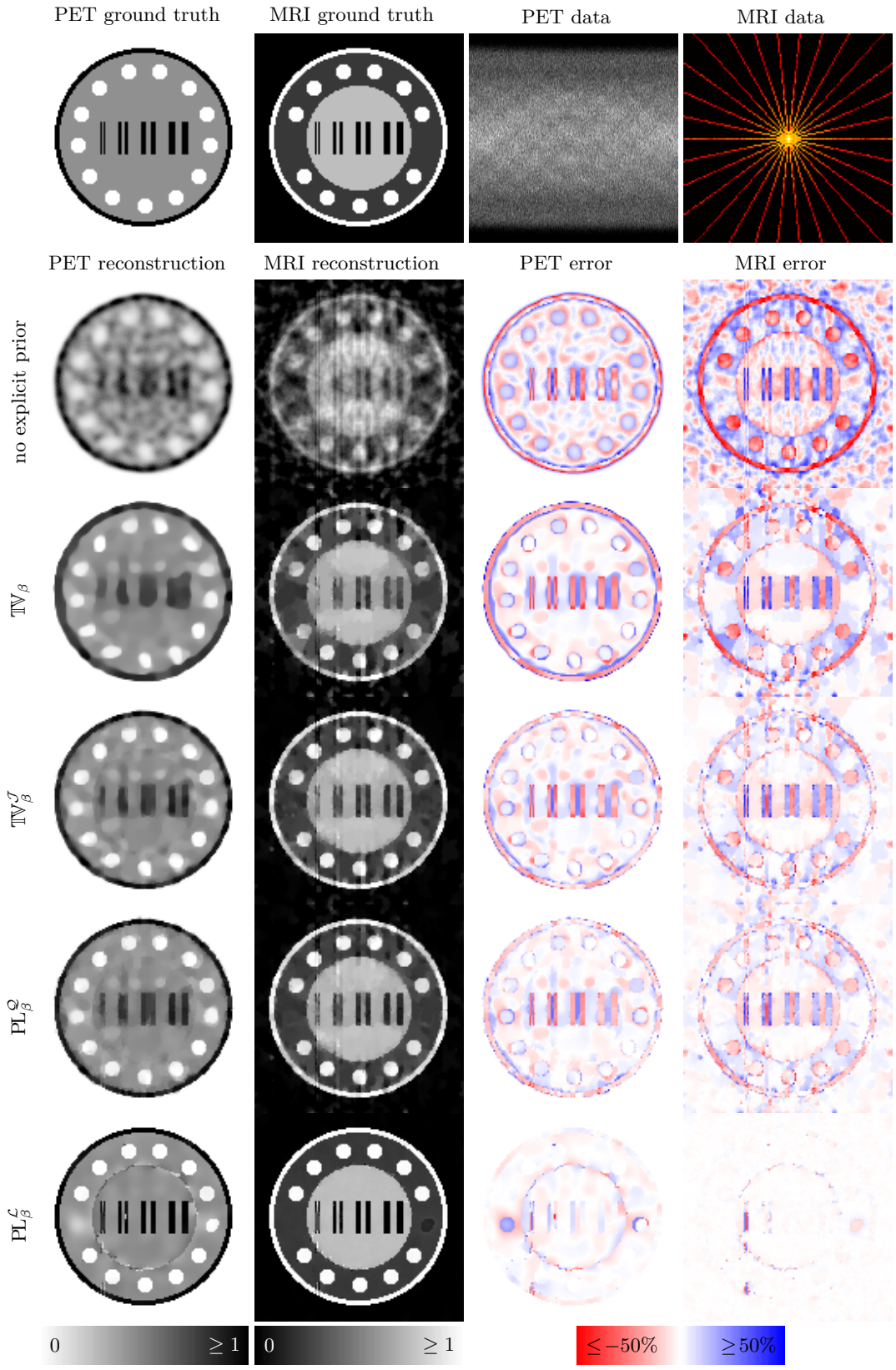


Figure 9.4. *Results for the PET-MRI data set* `radial15`*, i.e. MRI is sampled along 15 radial spokes, for early stopping / zero filling (no explicit prior), total variation ($\mathbb{TV}_\beta$), joint total variation ($\mathbb{TV}^{\mathcal{J}}_\beta$), quadratic parallel level sets ($\mathrm{PL}^{\mathcal{Q}}_\beta$) and linear parallel level sets ($\mathrm{PL}^{\mathcal{L}}_\beta$). Parameters are chosen to minimize the relative ℓ^2-error.*

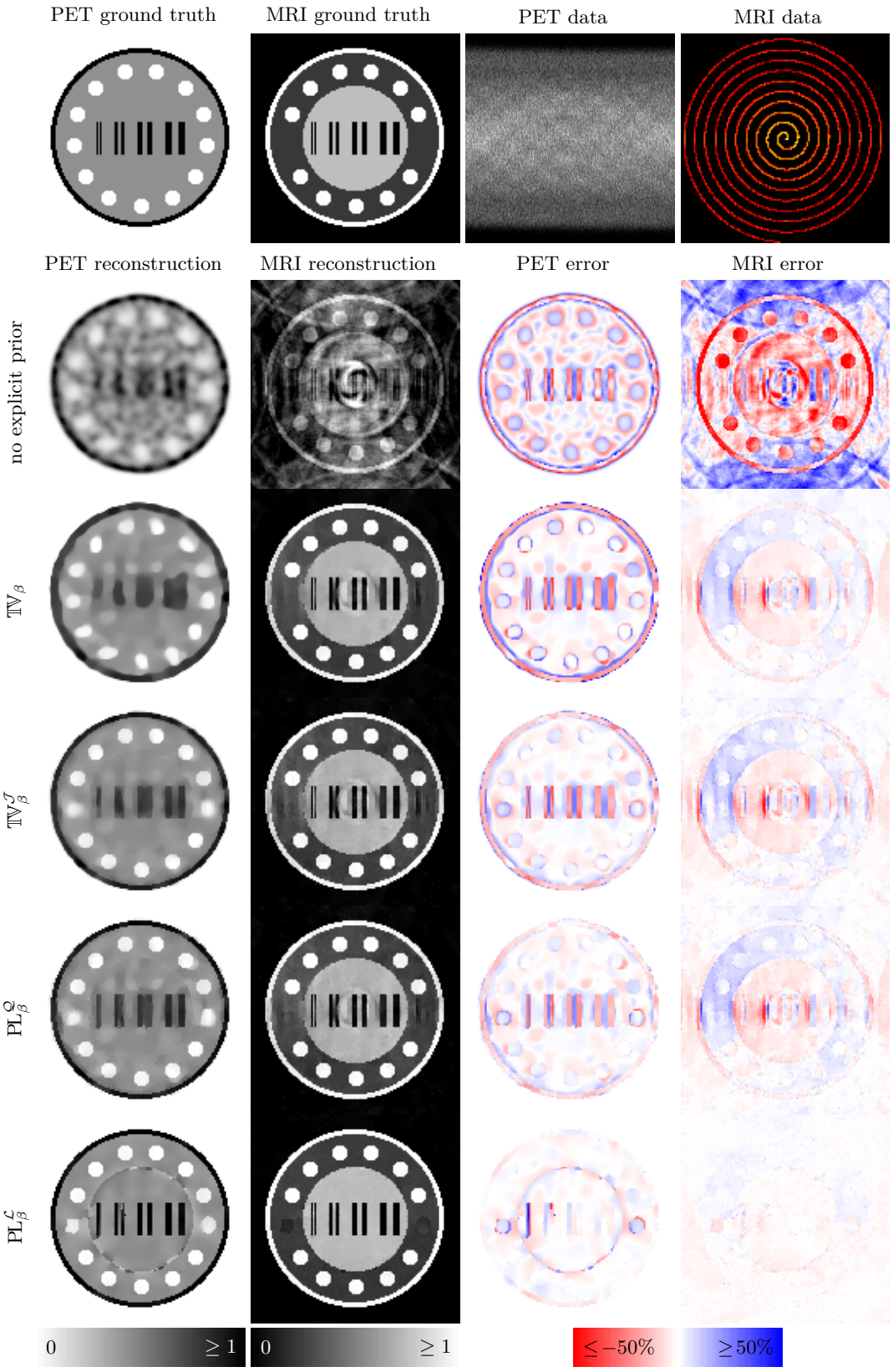


Figure 9.5. *Results for the PET-MRI data set* `spiralUni`*, i.e. MRI is sampled along a uniform spiral, for early stopping / zero filling (no explicit prior), total variation ($\mathbb{TV}_\beta$), joint total variation ($\mathbb{TV}^{\mathcal{J}}_\beta$), quadratic parallel level sets ($\mathrm{PL}^{\mathcal{Q}}_\beta$) and linear parallel level sets ($\mathrm{PL}^{\mathcal{L}}_\beta$). Parameters are chosen to minimize the relative ℓ^2-error.*

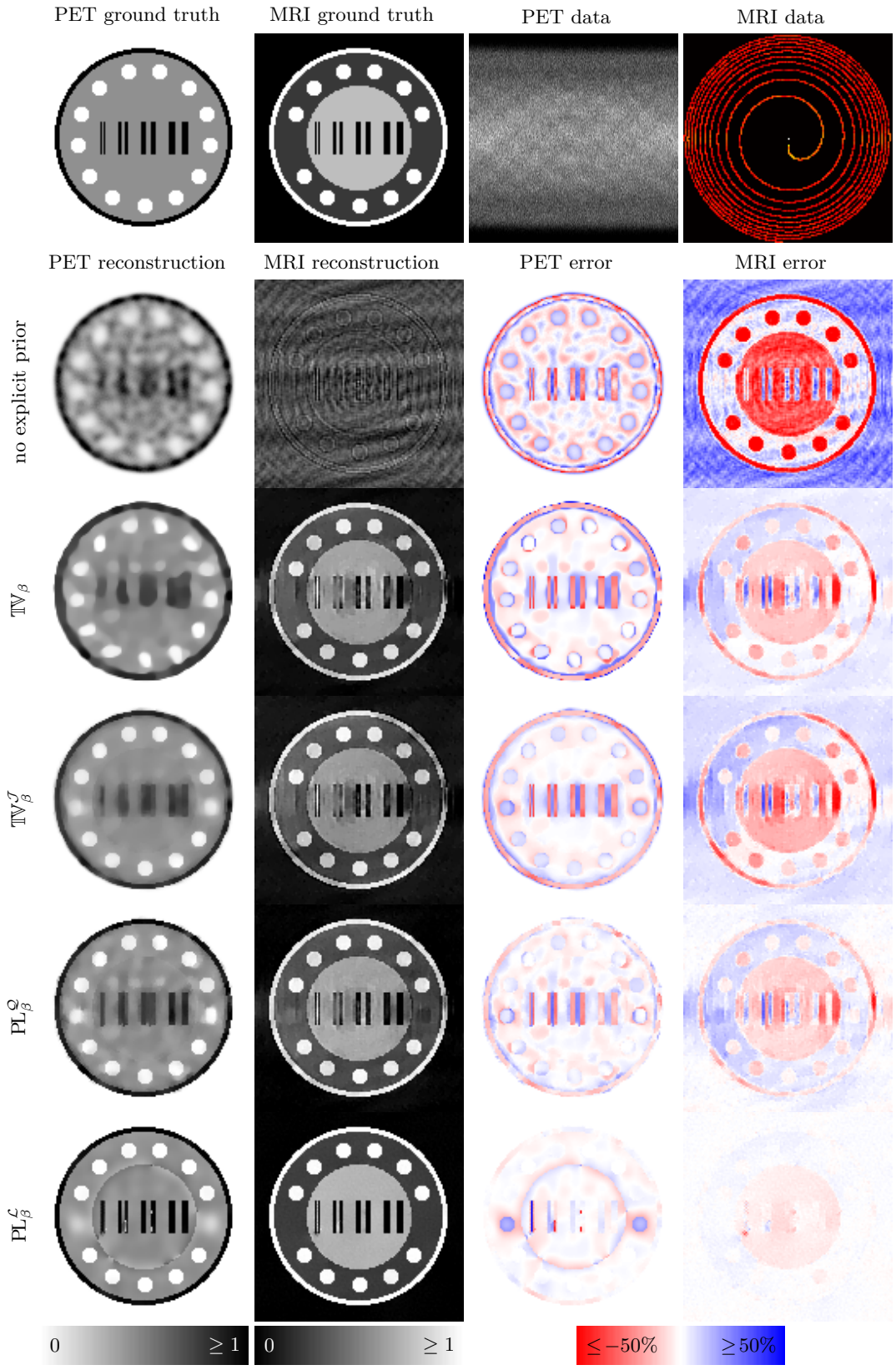


Figure 9.6. *Results for PET-MRI data set* `spiralHigh`*, i.e. MRI is fully sampled, for early stopping / zero filling (no explicit prior), total variation ($\mathbb{TV}_\beta$), joint total variation ($\mathbb{TV}^{\mathcal{J}}_\beta$), quadratic parallel level sets ($\mathrm{PL}^{\mathcal{Q}}_\beta$) and linear parallel level sets ($\mathrm{PL}^{\mathcal{L}}_\beta$). Parameters are chosen to minimize the relative ℓ^2-error.*

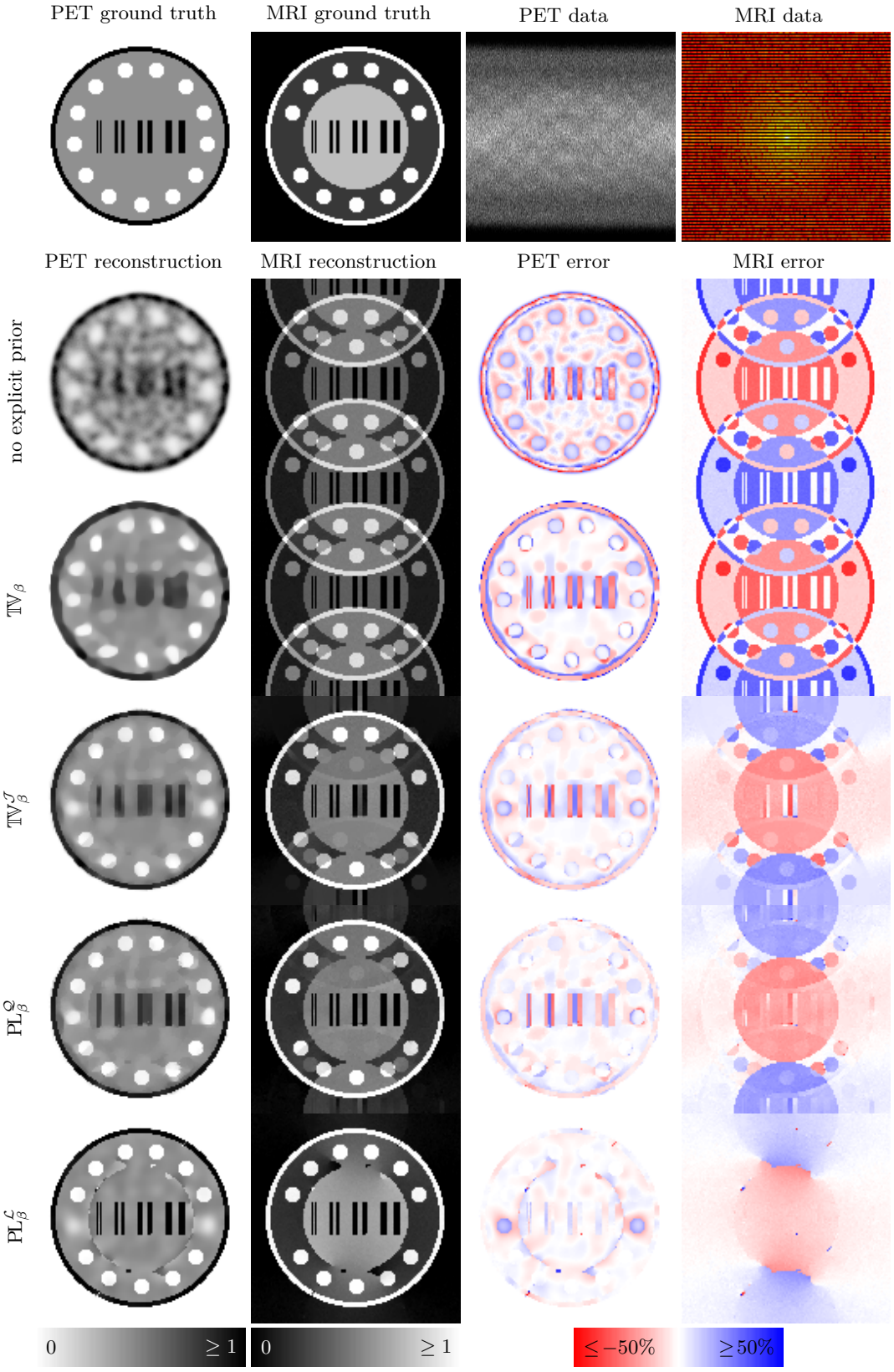


Figure 9.7. *Results for PET-MRI data set* `lines2`*, i.e. MRI is sampled at every other line, for early stopping / zero filling (no explicit prior), total variation ($\mathbb{TV}_\beta$), joint total variation ($\mathbb{TV}^{\mathcal{J}}_\beta$), quadratic parallel level sets ($\mathrm{PL}^{\mathcal{Q}}_\beta$) and linear parallel level sets ($\mathrm{PL}^{\mathcal{L}}_\beta$). Parameters are chosen to minimize the relative ℓ^2-error.*

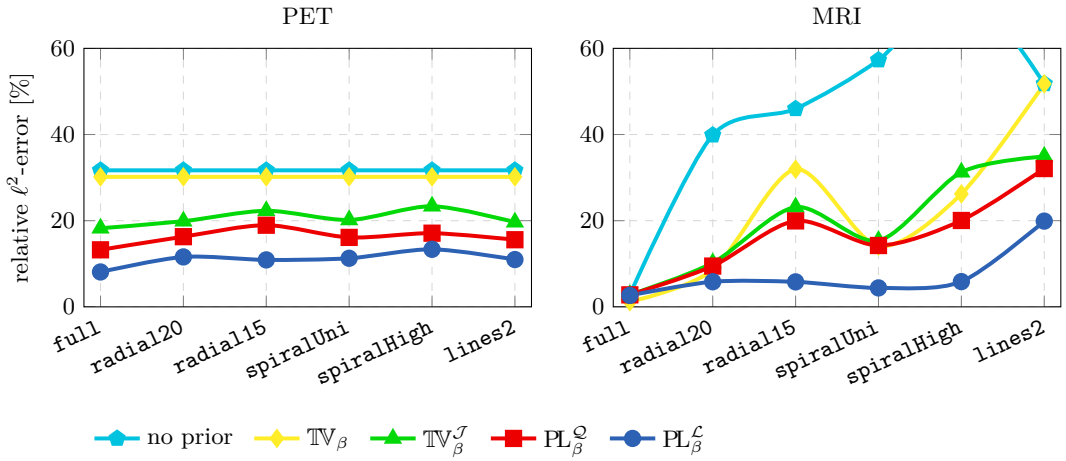


Figure 9.8. *Quantitative results for PET-MRI reconstruction with several MRI samplings and different reconstruction methods: no (explicit) prior, total variation ($\mathbb{TV}_\beta$), joint total variation ($\mathbb{TV}^{\mathcal{J}}_\beta$), quadratic parallel level sets ($\mathrm{PL}^{\mathcal{Q}}_\beta$) and linear parallel level sets ($\mathrm{PL}^{\mathcal{L}}_\beta$). The results for PET depends only on the MRI sampling in the case of joint reconstruction ($\mathbb{TV}^{\mathcal{J}}_\beta$,$\mathrm{PL}^{\mathcal{Q}}_\beta$, $\mathrm{PL}^{\mathcal{L}}_\beta$).*

Numerical Implementation. We minimize these functionals with L-BFGS-B [24, 117, 113] – a Quasi-Newton method with box constraints. The box constraints ensure that the images are always non-negative which is sufficient to guarantee that the evaluation of the Poisson likelihood is well-defined. An implementation of L-BFGS-B in Fortran is available on the web [25] and can be used in MATLAB with a mex wrapper [11]. The gradient of all these functionals can be computed as we have shown in section 3.5 for the data fidelities and 6.3 for the priors.

Data. We will present the results of two different test cases. In the first case, the ground truth is chosen to be a very simple circular phantom, cf. figure 9.2. The pixels are of size 1mm × 1mm for images with 128 × 128 pixels. The projections in PET are performed by an explicit represented matrix generated with NiftyRec [124] available at [122]. The resolution is modelled by a convolution in image space with a 2D Gaussian of full width at half maximum (FWHM) of 5mm in each dimension. The noise is controlled by scaling the phantom such that we expect to observe `1e+6` counts.

The second data set is a software brain phantom based on real data from [40] and converted to a continuous phantom using the software [71]. Thus, we can sample the phantoms at arbitrary resolution. For this experiment we have chosen again a pixel size of 1mm × 1mm but this time the phantom is of size 301 × 301 which is about 8 times larger in terms of the number of unknowns in the problem. The projections in PET for this phantom correspond to the geometry of one ring of the Siemens Biograph mMR and are computed using STIR [144] available online [143]. Additional to the intrinsic

resolution due to finite detector size, a Gaussian blur of 4mm × 4mm is employed. This time we also model attenuation, smoothly varying scatter and the randoms are chosen to be constant 1, see figure 9.1. In the case of the brain phantom, we simulated around 1.6e+6 counts.

In both cases, the MRI forward operator is modelled to be the discrete Fourier transform from which we sample with different sampling pattern. The noise is modelled to be Gaussian in both the real and imaginary part. It was scaled to have 4% and 10% of the total data energy in the k-space for the circular phantom and the brain phantom, respectively. We tried a variety of sampling pattern which are always plotted in the same figure at the top right corner.

Parameter Choice. We varied the overall regularization parameter α as well as the smoothing parameter β for all methods and chose the parameters that minimize the relative ℓ^2-error between the estimated image $w^\sharp$ and the ground truth w_0, i.e.

$$e_{\ell^2}(w^\sharp, w_0) \coloneqq \frac{|w^\sharp - w_0|}{|w_0|}, \tag{9.6}$$

where the images are treated as vectors to compute the norms.

The initial images for total variation and early stopping were chosen to be the zero image. For joint total variation and quadratic parallel level sets, we chose the initial guess to be the final result of total variation but only to speed up the reconstructions. In our experiments we observed that for quadratic parallel level sets – although it is not convex in the joint argument – the final reconstruction does not seem to depend on the initial guess. We used the result of quadratic parallel level sets as the input for linear parallel level sets as here the initial guess does matter. The side information for quadratic asymmetric parallel level sets was chosen to be the reconstruction with linear parallel level sets of the same modality. Although this strategy might be computationally prohibitive in practice, it shows the potential of this prior.

Evaluation. We will evaluate the results i) visually and ii) quantitatively in terms of the relative ℓ^2-error and SSIM [156].

9.1.2. Results for Simple Phantom. The visual results for a simple circular phantom are presented in figures 9.2 - 9.7. We will discuss here only the results in figure 9.4 in more detail. As it can be seen from the figure, for both PET and MRI the data is not

sufficient to reconstruct images of good visual quality when no explicit prior knowledge is used: In the case of PET, the main image features are reconstructed correctly but the reconstructed image is much blurrier than the ground truth. For MRI the reconstructed image without explicit prior information is severely degraded by undersampling artefacts. When total variation as a prior knowledge is incorporated – note that this is still separate reconstruction – especially the MRI reconstruction can be greatly improved. Nevertheless, there are still plenty of undersampling artefacts visible. Coupling the two reconstructions either with joint total variation or quadratic parallel level sets improves the PET reconstruction a lot and the MRI reconstruction a little bit. Using linear parallel level sets as a joint prior yields almost no errors any more in areas where both modalities share edge information. This comes at the price of smoothing unique features such as the blob on the left in PET and introducing fake edges.

The quantitative results shown in figure 9.8 are in accordance with the visual impression. The overall image quality in PET for separate reconstruction is in both cases almost the same independent of the kind of regularization. Introducing anatomical prior information from MRI enhances the PET reconstructions a lot whereby the two parallel level set priors seem to couple the modalities more strongly. These observations are almost independent of the undersampling we employed for MRI. The quantitative results for MRI show that for this example in every case of undersampling, coupling the two inverse problems with linear parallel level sets results in superior reconstructions compared to separate reconstructions and joint reconstructions with the other two priors.

9.1.3. Results for Brain Phantom. The visual results for the brain phantom are shown in figures 9.9 - 9.14. Many observations for this test case are the same as in the case of the simple phantom. Let us note only a few observations. First, for all cases of sampling, coupling the modalities with linear parallel level sets yields very well defined structures in the PET reconstructions. Second, this time only in figure 9.14 can we observe that the symmetric parallel level sets prior improves the MRI reconstruction. Finally, note that the quadratic asymmetric parallel level set prior yields the best results in all cases of sampling.

As before, our visual impression are underlined by the quantitative results in figure 9.15. Joint reconstruction improves the results for PET in every cases of undersampling. This time only in two cases of undersampling are the MRI reconstruction significantly

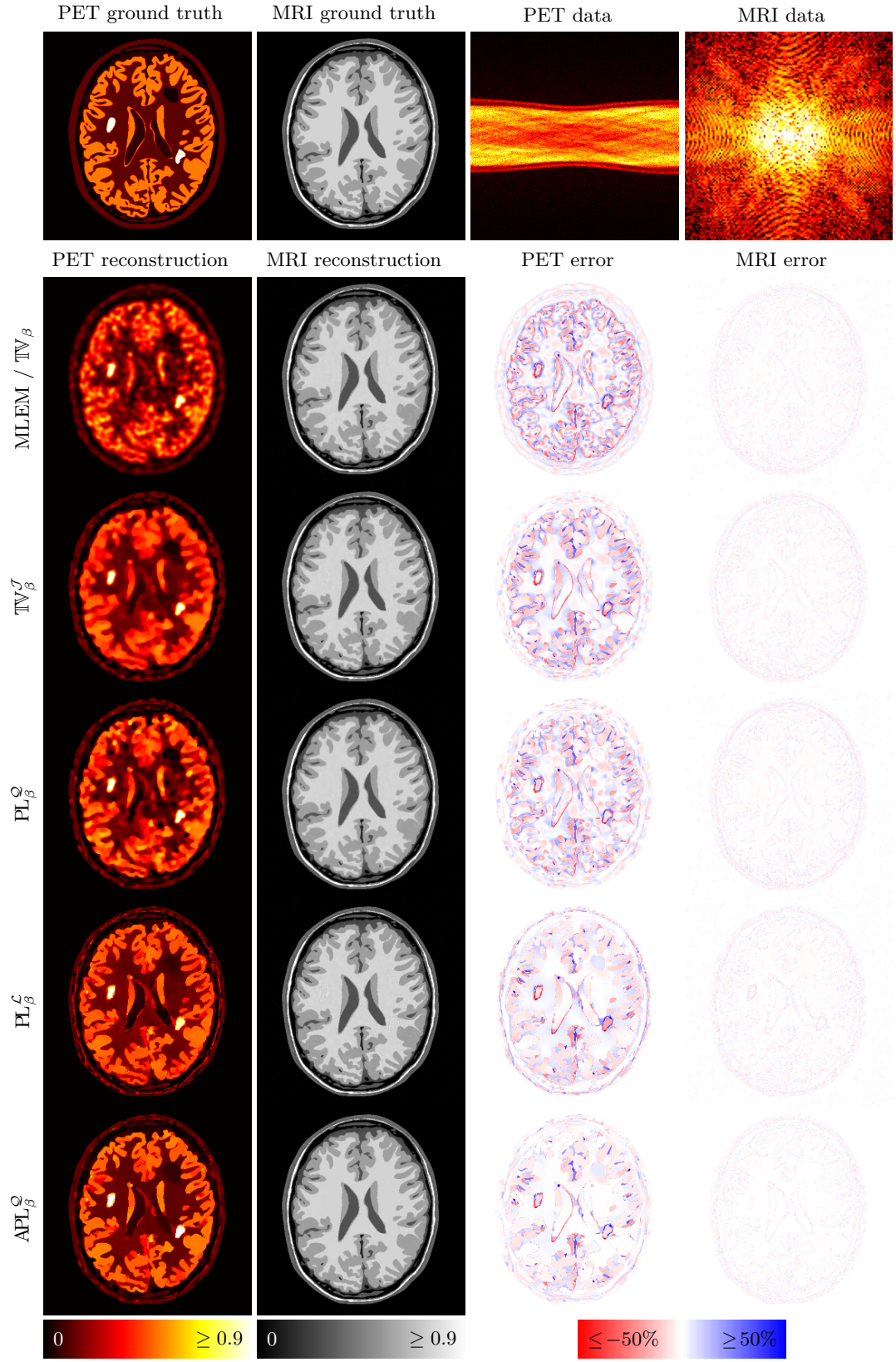


Figure 9.9. *Results for PET-MRI data set* `full`*, i.e. MRI is fully sampled, for MLEM with early stopping, total variation ($\mathbb{TV}_\beta$), joint total variation ($\mathbb{TV}^{\mathcal{J}}_\beta$), quadratic parallel level sets ($\mathrm{PL}^{\mathcal{Q}}_\beta$), linear parallel level sets ($\mathrm{PL}^{\mathcal{L}}_\beta$) and asymmetric quadratic parallel level sets ($\mathrm{APL}^{\mathcal{Q}}_\beta$). Parameters are chosen to minimize the relative ℓ^2-error.*

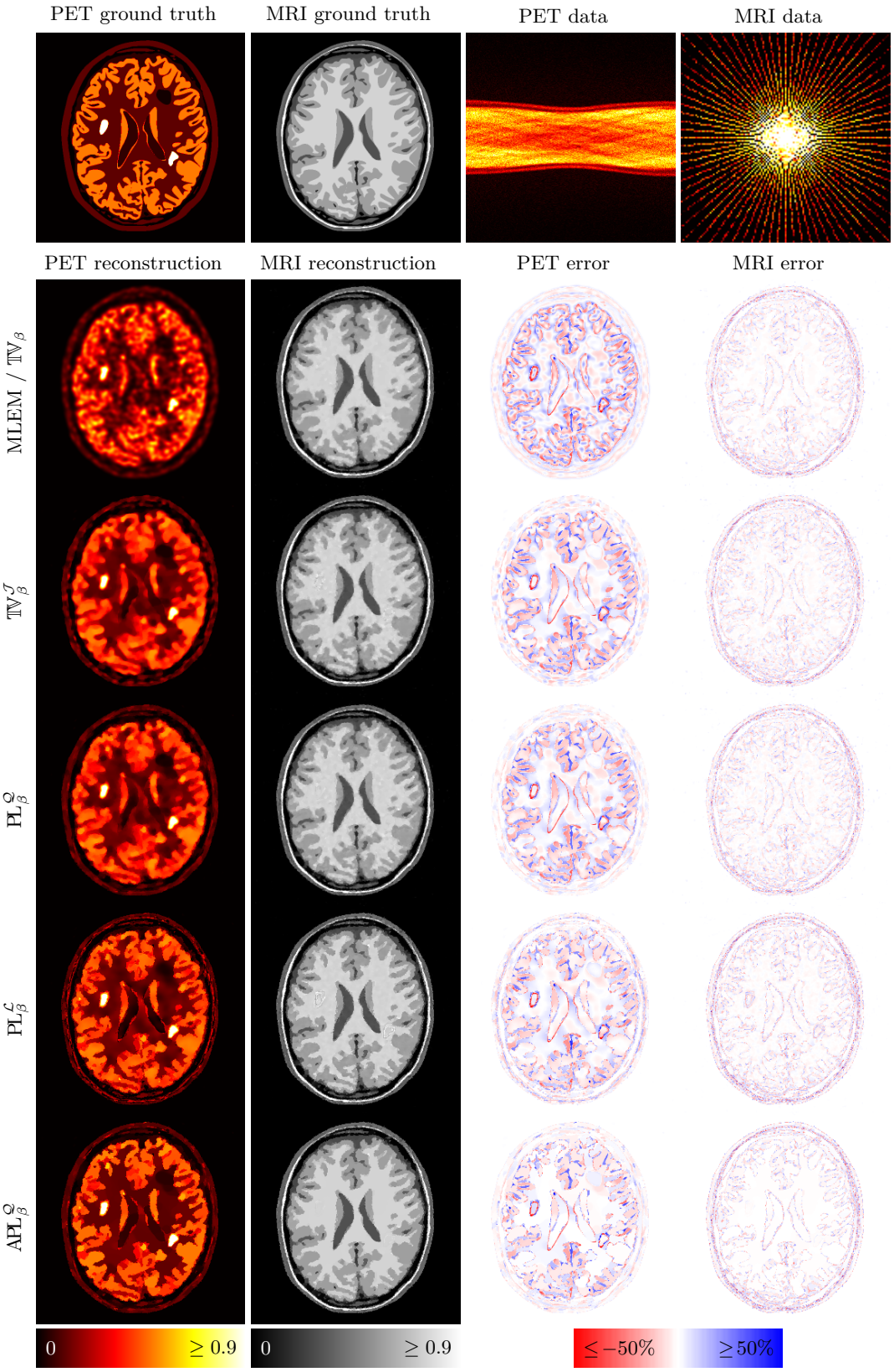


Figure 9.10. *Results for PET-MRI data set* `radial40`*, i.e. MRI is sampled along forty radial spokes, for MLEM with early stopping, total variation ($\mathbb{TV}_\beta$), joint total variation ($\mathbb{TV}^{\mathcal{J}}_\beta$), quadratic parallel level sets ($\mathrm{PL}^{\mathcal{Q}}_\beta$), linear parallel level sets ($\mathrm{PL}^{\mathcal{L}}_\beta$) and asymmetric quadratic parallel level sets ($\mathrm{APL}^{\mathcal{Q}}_\beta$). Parameters are chosen to minimize the relative ℓ^2-error.*

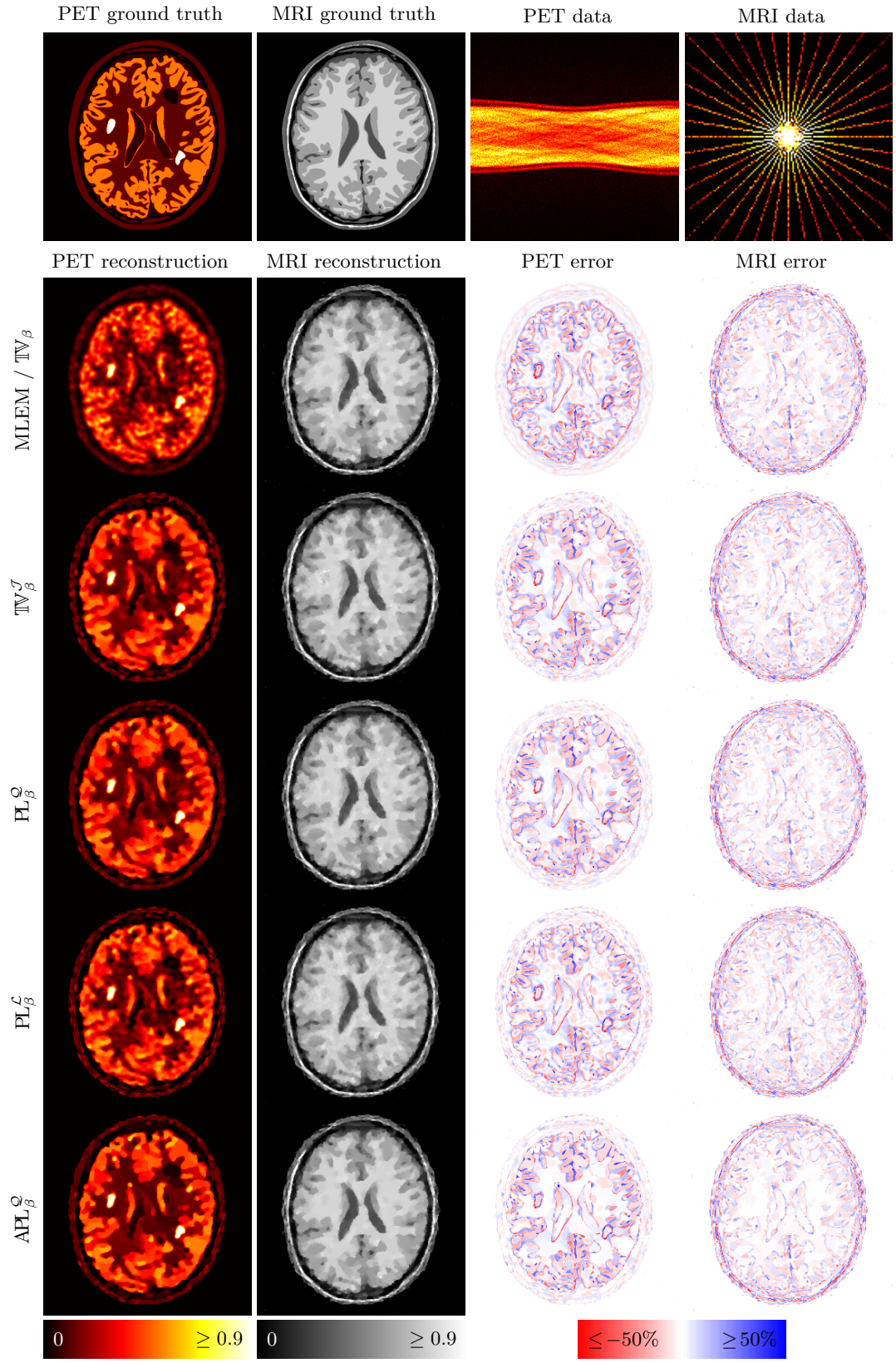


Figure 9.11. *Results for PET-MRI data set* `radial20`*, i.e. MRI is sampled along twenty radial spokes, for MLEM with early stopping, total variation ($\mathbb{TV}_\beta$), joint total variation ($\mathbb{TV}^{\mathcal{J}}_\beta$), quadratic parallel level sets ($\mathrm{PL}^{\mathcal{Q}}_\beta$), linear parallel level sets ($\mathrm{PL}^{\mathcal{L}}_\beta$) and asymmetric quadratic parallel level sets ($\mathrm{APL}^{\mathcal{Q}}_\beta$). Parameters are chosen to minimize the relative ℓ^2-error.*

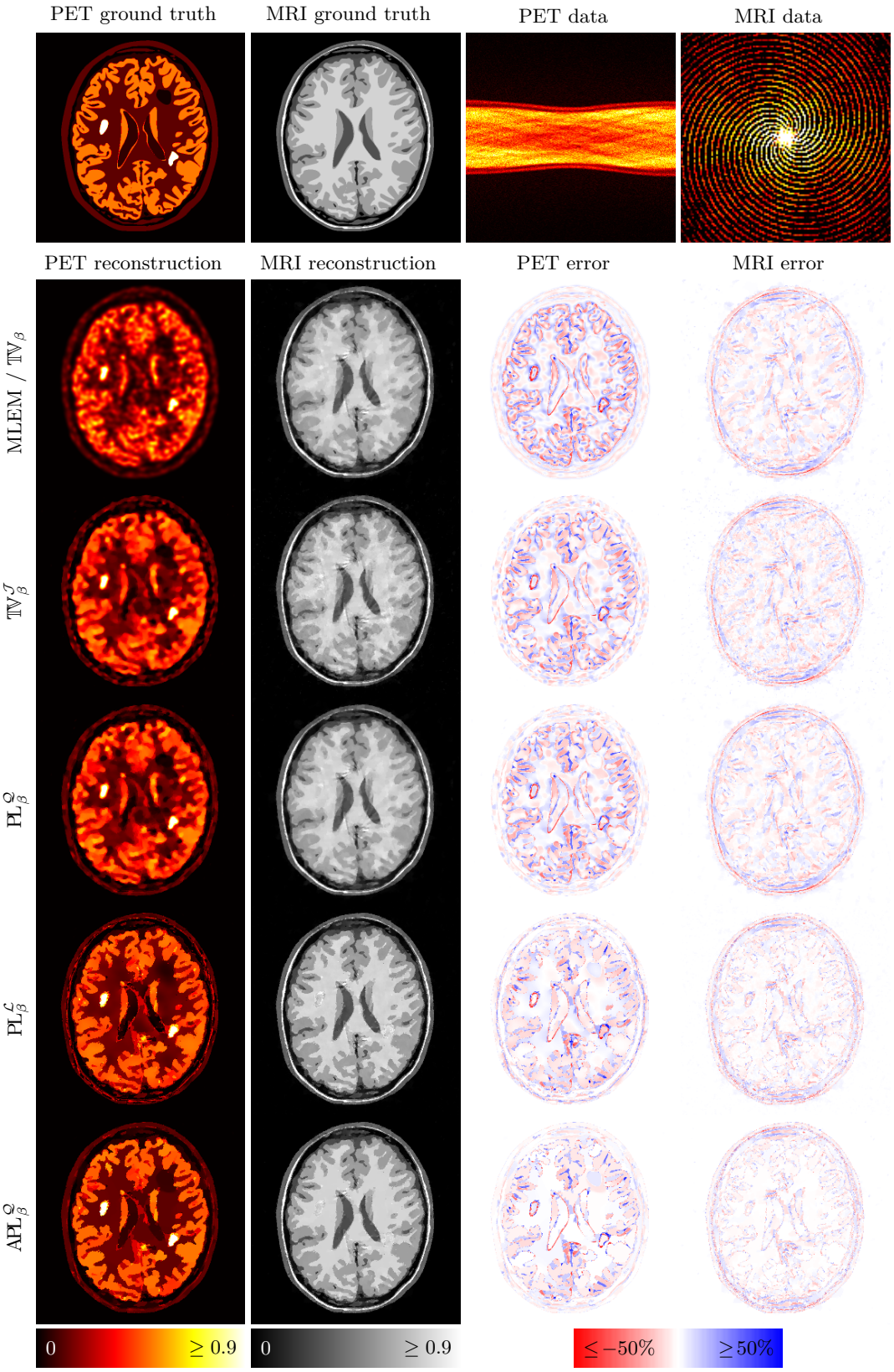


Figure 9.12. *Results for PET-MRI data set* `spiralUni20`*, i.e. MRI is sampled along twenty uniform spirals, for MLEM with early stopping, total variation ($\mathbb{TV}_\beta$), joint total variation ($\mathbb{TV}^{\mathcal{J}}_\beta$), quadratic parallel level sets ($\mathrm{PL}^{\mathcal{Q}}_\beta$), linear parallel level sets ($\mathrm{PL}^{\mathcal{L}}_\beta$) and asymmetric quadratic parallel level sets ($\mathrm{APL}^{\mathcal{Q}}_\beta$). Parameters are chosen to minimize the relative ℓ^2-error.*

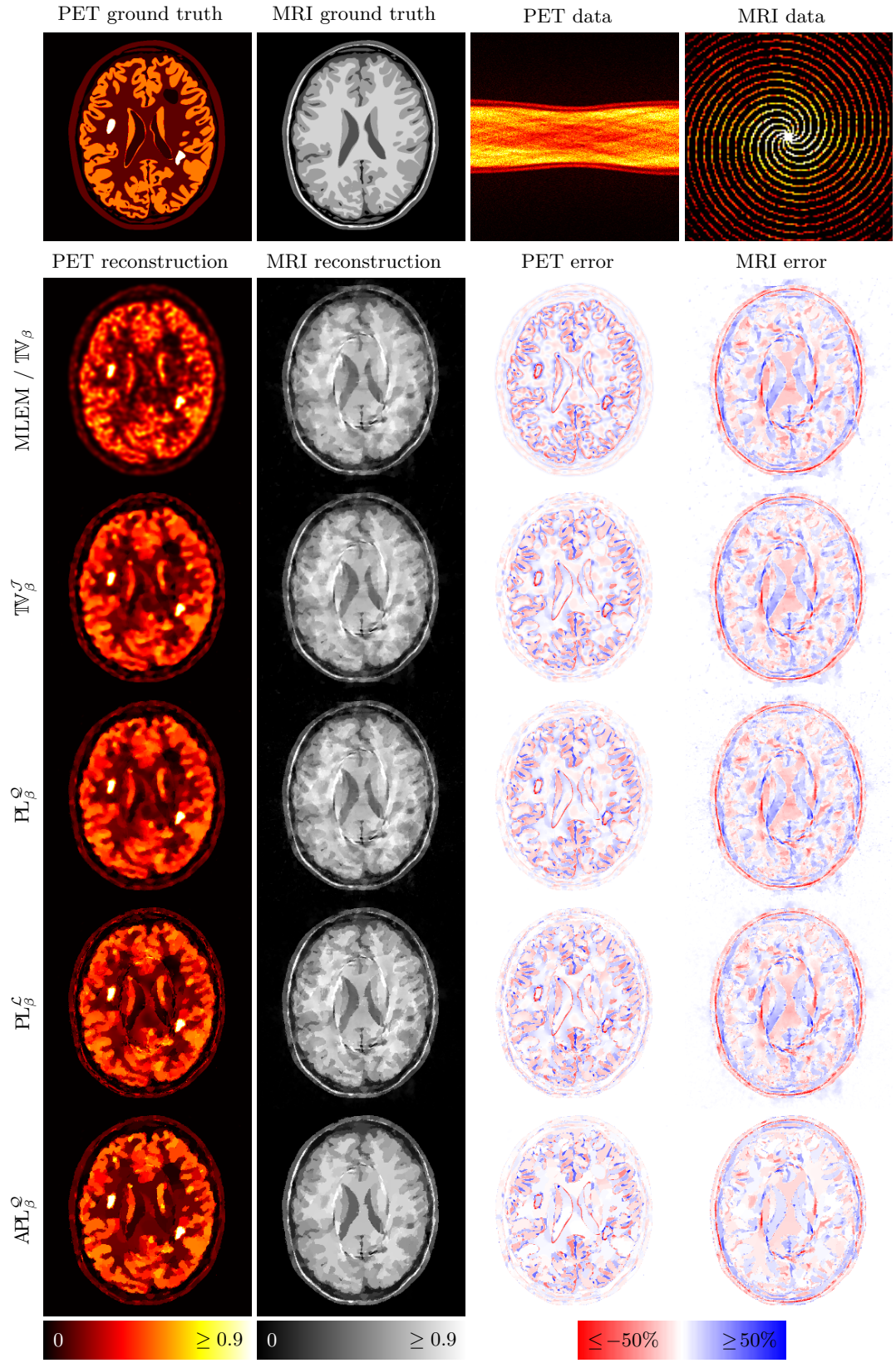


Figure 9.13. *Results for PET-MRI data set* `spiralUni10`*, i.e. MRI is sampled along ten uniform spirals, for MLEM with early stopping, total variation ($\mathbb{TV}_\beta$), joint total variation ($\mathbb{TV}^{\mathcal{J}}_\beta$), quadratic parallel level sets ($\mathrm{PL}^{\mathcal{Q}}_\beta$), linear parallel level sets ($\mathrm{PL}^{\mathcal{L}}_\beta$) and asymmetric quadratic parallel level sets ($\mathrm{APL}^{\mathcal{Q}}_\beta$). Parameters are chosen to minimize the relative ℓ^2-error.*

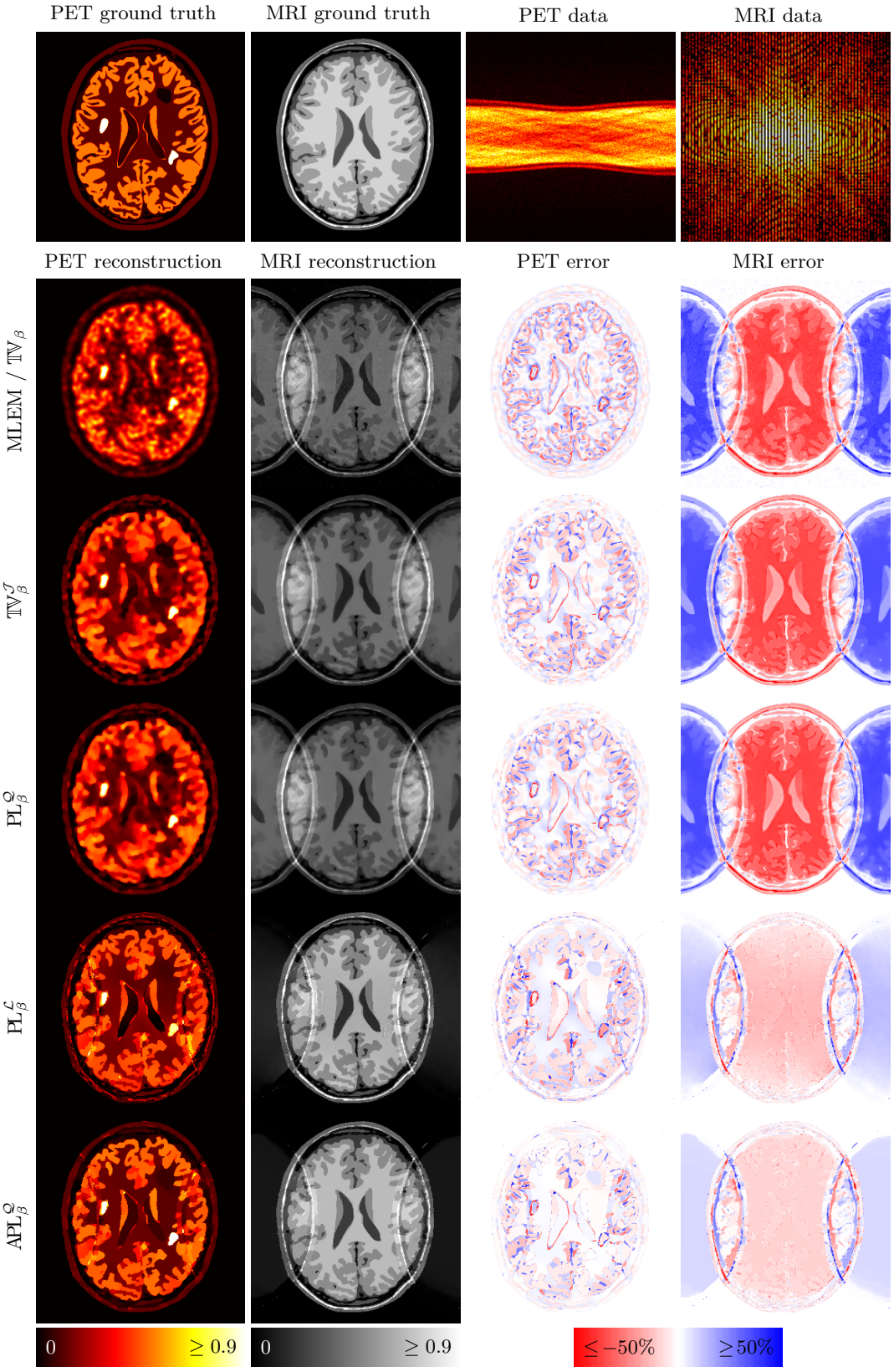


Figure 9.14. *Results for PET-MRI data set* `lines2`*, i.e. MRI is sampled at every other line, for MLEM with early stopping, total variation ($\mathbb{TV}_\beta$), joint total variation ($\mathbb{TV}^{\mathcal{J}}_\beta$), quadratic parallel level sets ($\mathrm{PL}^{\mathcal{Q}}_\beta$), linear parallel level sets ($\mathrm{PL}^{\mathcal{L}}_\beta$) and asymmetric quadratic parallel level sets ($\mathrm{APL}^{\mathcal{Q}}_\beta$). Parameters are chosen to minimize the relative relative ℓ^2-error.*

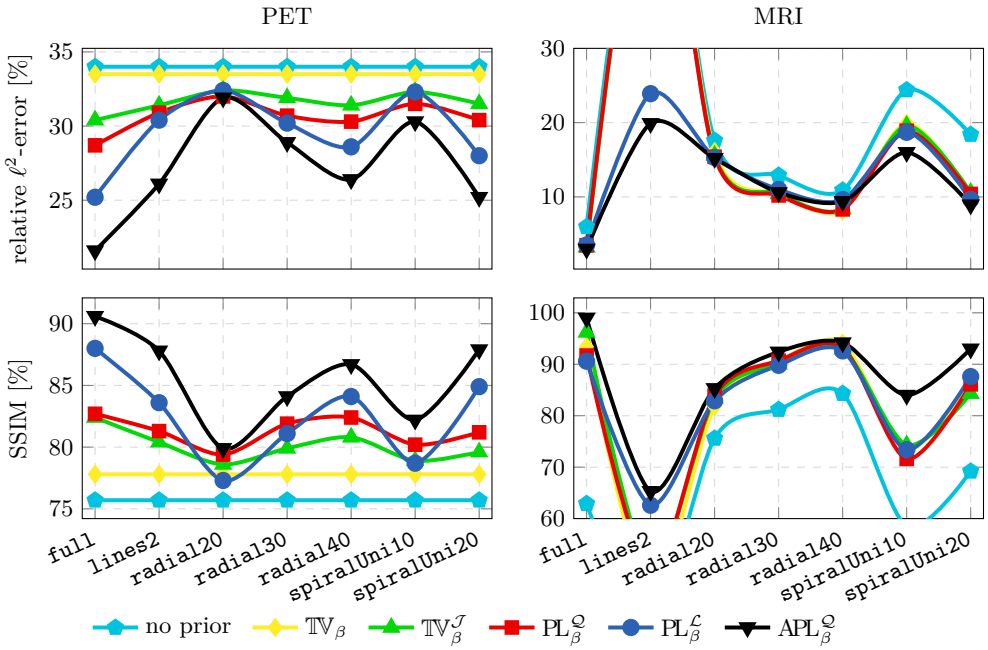


Figure 9.15. *Quantitative results of joint reconstruction for PET-MRI with the brain phantom, different kind of MRI samplings and different methods: no explicit prior, total variation ($\mathbb{TV}_\beta$), joint total variation ($\mathbb{TV}^{\mathcal{J}}_\beta$), quadratic parallel level sets ($\mathrm{PL}^{\mathcal{Q}}_\beta$) and linear parallel level sets ($\mathrm{PL}^{\mathcal{L}}_\beta$). The results for PET depends only on the MRI sampling in the case of joint reconstruction.*

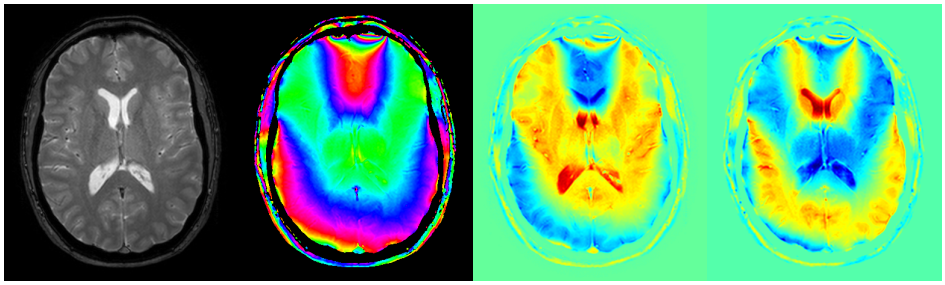


Figure 9.16. *From left to right: magnitude, phase, real part and imaginary part of a real (non-simulated) MRI image obtained from a gradient echo sequence acquired on a Philips 3T Achieva scanner. It is clear that different priori information should be used on each of these images. Images courtesy of UCL hospital.*

better for joint reconstruction than for separate reconstruction. It is interesting to see that although the MRI reconstructions for `lines2` are the worst, the corresponding PET reconstructions are as good as in the case of `radial40` and `spiralUni20`.

9.2. Phase in MRI and its Implications on Prior Knowledge.

Up to this point we always assumed that the images we want to reconstruct in MRI are real-valued. However, in practice due to model imperfections a more accurate model would include phase as well. The phase is often considered to be an unwanted by-product but has recently drawn a lot of attention as in some applications it is valuable itself [138]. An MRI image with magnitude and phase is shown in figure 9.16.

To make use of compressed sensing for undersampled MRI reconstruction, we have to include a priori knowledge on the images we seek. From figure 9.16 we can see that the magnitude image shows distinct boundaries with little variation within these areas and the phase changes smoothly almost everywhere where it is properly defined, i.e. the magnitude being strictly positive. The combination of the two yields a complex image with its real and imaginary parts also shown in figure 9.16 on the right. While the phase and magnitude are quite regular, it is difficult to define the structure in the real and imaginary part.

Although MRI images are very often considered as complex numbers, the a priori information is usually motivated by magnitude images only. Nevertheless, many reconstruction methods using priors are then implemented with complex numbers. We will show in this section what total variation of a complex image means and relate it to another prior we have seen already. Better modelling of the a priori knowledge is based on a different parametrization in terms of magnitude and phase and we analyse the resulting new data term. The numerical results show that with this modification images with a lot less artefacts can be reconstructed so that the acquisition time can be further reduced.

Before we present our results we would like to cite a comment by Michael Lustig – the pioneer in compressed sensing for MRI – about phase in MRI:

> *In MRI, instrumental sources of phase errors can cause low-order phase variation. These carry no physical information, but create artificial variation in the image which makes it more difficult to sparsify, especially by finite differences. By estimating the phase variation, the reconstruction can be significantly improved. This phase estimate may be obtained using very low-resolution fully sampled k-space information.* (Lustig [103])

9.2.1. Complex Data Fidelity. To apply gradient-based optimization techniques we have to compute the derivatives of the MRI data fidelity

$$\mathcal{E}(Bv|g) = \frac{1}{2\sigma^2} |Bv - g|^2 \tag{9.7}$$

with respect to the real and imaginary part.

Lemma 9.2.1. *Let* $B\colon \mathbb{C}^N \to \mathbb{C}^M$ *be linear. The derivatives of* $\mathcal{E}\colon \mathbb{R}^{2N} \simeq \mathbb{C}^N \to \mathbb{R}$

defined by (9.7) *with respect to the real part* v_1 *and imaginary part* v_2 *are given by*

$$\partial_{v_1}\mathcal{E}(Bv|g) = \frac{1}{\sigma^2}\operatorname{Re}\left[B^*(Bv-g)\right] \tag{9.8}$$

$$\partial_{v_2}\mathcal{E}(Bv|g) = \frac{1}{\sigma^2}\operatorname{Im}\left[B^*(Bv-g)\right]. \tag{9.9}$$

Proof. Using $v = \operatorname{Re}^* v_1 + \operatorname{Im}^* v_2$, there is

$$\mathcal{E}(Bv|g) = \frac{1}{2\sigma^2}\left|B(\operatorname{Re}^* v_1 + \operatorname{Im}^* v_2) - g\right|^2 \tag{9.10}$$

$$= \frac{1}{2\sigma^2}\left|(B\operatorname{Re}^*)v_1 + (B\operatorname{Im}^* v_2 - g)\right|^2 \tag{9.11}$$

$$= \frac{1}{2\sigma^2}\left|(B\operatorname{Im}^*)v_2 + (B\operatorname{Re}^* v_1 - g)\right|^2. \tag{9.12}$$

We can now apply usual rules to calculate the derivatives of real vectors and derive

$$\partial_{v_1}\mathcal{E}(Bv|g) = \frac{1}{\sigma^2}(B\operatorname{Re}^*)^*[(B\operatorname{Re}^*)v_1 + (B\operatorname{Im}^* v_2 - g)] = \frac{1}{\sigma^2}\operatorname{Re}\left[B^*(Bv-g)\right] \tag{9.13}$$

and

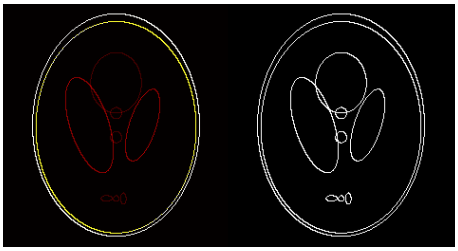
$$\partial_{v_2}\mathcal{E}(Bv|g) = \frac{1}{\sigma^2}\operatorname{Im}\left[B^*(Bv-g)\right]. \tag{9.14}$$

■

9.2.2. Total Variation on Complex Images. A common choice of prior knowledge in MRI is that the image is sparse in its gradient domain, i.e. its spatial gradient (or finite differences) is zero at many locations. Usually, this is motivated by the visual appearance of magnitude MRI images, e.g. figure 9.16, but nevertheless the method is applied for complex images. The pitfall here is that the algorithms used for these kind of problems are usually well-defined and convergent for both real and complex images but as we will see in the numerical experiments, might lead to inferior results due to the phase. The complex image can be arbitrarily non-sparse in the gradient domain independent of the gradient-sparsity of the magnitude image, see figure 9.17. We will now discuss what total variation for complex images actually does.

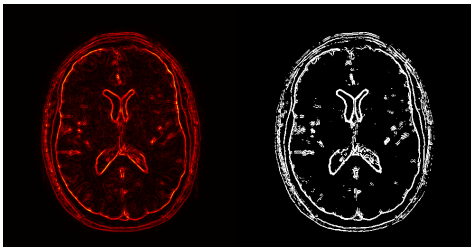
If we denote our complex MRI image v with real and imaginary part v_1, v_2, then at each position x the gradient can be computed as $\nabla v(x) = \nabla v_1(x) + \mathtt{i}\nabla v_2(x)$ and its norm is given by $|\nabla v(x)|^2 = |\nabla v_1(x)|^2 + |\nabla v_2(x)|^2$. Then the total variation of v can

(a) phantom: magnitude, phase and wrapped phase

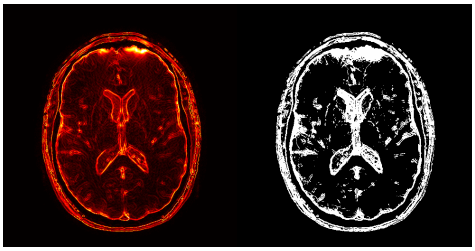


(b) gradient of magnitude

(c) gradient complex image



(d) gradient of magnitude



(e) gradient of complex image

Figure 9.17. *Phase in MRI ruins sparsity. Each sub-figure shows the norm (left) and the support (right) of the gradient of either a magnitude or a complex image. Although the magnitude image is sparse in the gradient domain (b), the complex image is not when there are large phase variations (c). This can also be seen for the data from figure 9.16 where the magnitude (d) is much sparser than the complex image (e). The colour scales are [0,100] except for the wrapped phase [0,2π]. The support is defined to be larger than 1 (b,c) and 15 (d,e) for pixels of unit size.*

be written as

$$\mathbb{TV}(v) = \int |\nabla v(x)| \, \mathrm{d}x = \int \left(|\nabla v_1(x)|^2 + |\nabla v_2(x)|^2\right)^{1/2} \mathrm{d}x = \mathbb{TV}^{\mathcal{J}}(v_1, v_2). \tag{9.15}$$

Thinking of v not as a complex but a vector-valued image (with two real-valued components) then we can see from (9.15) that total variation on a complex image actually corresponds to joint total variation on its real and imaginary part. It is known that this prior favours images with spatial gradients at similar locations.

Total Variation on Magnitude. An option to circumvent this problem is to formulate the total variation prior on the magnitude image. A naive way of dealing with the problem of having knowledge on the magnitude but using the usual linear parametrization with the real and imaginary part is to phrase the prior on the magnitude. For a complex MRI image v with real and imaginary part v_1, v_2, the magnitude is given by

$r \coloneqq |v| = (v_1^2 + v_2^2)^{1/2}$. In case we expect the magnitude r to be piecewise constant, then we could equivalently expect $r^2 = v_1^2 + v_2^2$ to be piecewise constant. The total variation on this quantity is

$$\mathbb{TV}(r^2) = \int_\Omega |\nabla(v_1^2 + v_2^2)| = 2 \int_\Omega |v_1 \nabla v_1 + v_2 \nabla v_2|. \tag{9.16}$$

In this functional, image gradients are weighted with image values. This leads to the undesirable property that gradients in areas of small image values are less penalized than gradients in areas of large values. Note also that this form of total variation on the magnitude is non-convex.

9.2.3. Non-Linear Parametrisation: Magnitude and Phase. A third option to solve this dilemma is to change the underlying parametrization of the complex image. Instead of real and imaginary part, we can parametrize the complex image in terms of its magnitude and phase [61, 165, 164, 150]. Recall that the phase for a complex number v is a number $p \in \mathbb{R}$ such that $v = r \exp(\mathtt{i}p)$. To be uniquely defined we can constrain the phase to any right-closed and left-open interval of length 2π such as $[0, 2\pi)$.

With the polar-to-Cartesian mapping $M \colon [0, \infty)^N \times [0, 2\pi)^N \to \mathbb{C}^N$,

$$M(r, p) \coloneqq r \exp(\mathtt{i}p) \tag{9.17}$$

we can write the non-linear MRI forward operator as $E \coloneqq B \circ M$ and the data fidelity term becomes

$$\mathcal{E}(E(r, p)|g) = \frac{1}{2\sigma^2} |E(r, p) - g|^2 . \tag{9.18}$$

With this parametrization we can specify our prior beliefs directly on the magnitude and phase.

$$\mathcal{J}(r, p) = \mathcal{E}(E(r, p)|g) + \alpha_1 \mathcal{P}_1(r) + \alpha_2 \mathcal{P}_2(p) \tag{9.19}$$

Unfortunately, this comes at the cost of making the MRI forward operator *non-linear* therefore possibly non-convex, cf. figure 9.18.

For gradient-based optimization, we need to compute the derivatives of this data fidelity with respect to magnitude and phase.

Lemma 9.2.2. *Let* $\Phi \coloneqq \operatorname{diag}(\exp(\mathtt{i}p))$ *and* $R \coloneqq \operatorname{diag}(r)$. *Note that* $\Phi^* = \Phi^{-1} = \operatorname{diag}(\exp(-\mathtt{i}p))$. *Then the derivatives of* $\mathcal{E}\colon \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ *defined as* (9.18) *are given by*

$$\partial_r \mathcal{E}(E(r,p)|g) = \frac{1}{\sigma^2} \operatorname{Re}\{\Phi^{-1} B^* [E(r,p) - g]\} \tag{9.20}$$

$$\partial_p \mathcal{E}(E(r,p)|g) = \frac{1}{\sigma^2} R \operatorname{Im}\{\Phi^{-1} B^* [E(r,p) - g]\}. \tag{9.21}$$

Proof. First of all, note that since

$$E(r,p)_m = \sum_{k=1}^{K} B_{m,k} r_k \exp(\mathtt{i}p_k) \tag{9.22}$$

$$\begin{aligned} &= \sum_{k=1}^{K} r_k \big[\operatorname{Re} B_{m,k} \cos p_k - \operatorname{Im} B_{m,k} \sin p_k \\ &\qquad + \mathtt{i}(\operatorname{Re} B_{m,k} \sin p_k + \operatorname{Im} B_{m,k} \cos p_k)\big] \end{aligned} \tag{9.23}$$

there is

$$\partial_{r_n} E(r,p)_m = B_{m,n} \exp(\mathtt{i}p_n) \tag{9.24}$$

$$\begin{aligned} \partial_{p_n} E(r,p)_m &= \mathtt{i} r_n \big[\operatorname{Re} B_{m,n} \cos p_n - \operatorname{Im} B_{m,n} \sin p_n \\ &\qquad + \mathtt{i}(\operatorname{Re} B_{m,n} \sin p_n + \operatorname{Im} B_{m,n} \cos p_n)\big] \end{aligned} \tag{9.25}$$

$$= \mathtt{i} r_n B_{m,n} \exp(\mathtt{i}p_n) \tag{9.26}$$

Note that

$$\operatorname{Re}[\partial_{p_n} E(r,p)_m] = \operatorname{Im}[\overline{r_n B_{m,n} \exp(\mathtt{i}p_n)}] \tag{9.27}$$

$$\operatorname{Im}[\partial_{p_n} E(r,p)_m] = \operatorname{Re}[\overline{r_n B_{m,n} \exp(\mathtt{i}p_n)}]. \tag{9.28}$$

By definition of (9.18) we have

$$\mathcal{E}(E(r,p)|g) = \frac{1}{2\sigma^2} \sum_{m=1}^{M} |E(r,p)_m - g_m|^2 \tag{9.29}$$

$$= \frac{1}{2\sigma^2} \sum_{m=1}^{M} \{[\operatorname{Re}(E(r,p) - g)_m]^2 + [\operatorname{Im}(E(r,p) - g)_m]^2\}. \tag{9.30}$$

As derivatives and Re, Im commute, i.e. $\partial_x \operatorname{Re} y = \operatorname{Re} \partial_x y$ and $\partial_x \operatorname{Im} y = \operatorname{Im} \partial_x y$ there is

$$\partial_x \mathcal{E}(E(r,p)|g) = \frac{1}{\sigma^2} \sum_{m=1}^{M} \{[\operatorname{Re}(E(r,p)-g)_m] \operatorname{Re}[\partial_x (E(r,p))_m] + [\operatorname{Im}(E(r,p)-g)_m] \operatorname{Im}[\partial_x (E(r,p))_m]\}. \tag{9.31}$$

Then there is

$$\begin{aligned}
\partial_{r_n} \mathcal{E}(E(r,p)|g) &\overset{(9.31)}{=} \frac{1}{\sigma^2} \sum_{m=1}^{M} \{[\operatorname{Re}(E(r,p)-g)_m] \operatorname{Re}[\partial_{r_n} (E(r,p))_m] \\
&\quad + [\operatorname{Im}(E(r,p)-g)_m] \operatorname{Im}[\partial_{r_n} (E(r,p))_m]\} && (9.32) \\
&\overset{(9.24)}{=} \frac{1}{\sigma^2} \sum_{m=1}^{M} \{[\operatorname{Re}(E(r,p)-g)_m] \operatorname{Re}[B_{m,n} \exp(\mathtt{i} p_n)] \\
&\quad + [\operatorname{Im}(E(r,p)-g)_m] \operatorname{Im}[B_{m,n} \exp(\mathtt{i} p_n)]\} && (9.33) \\
&= \operatorname{Re}\{(B\Phi)^* [E(r,p)-g]\}_n = \operatorname{Re}\{\Phi^{-1} B^* [E(r,p)-g]\}_n && (9.34)
\end{aligned}$$

and therefore

$$\begin{aligned}
\partial_{p_n} \mathcal{E}(E(r,p)|g) &\overset{(9.31)}{=} \frac{1}{\sigma^2} \sum_{m=1}^{M} \{[\operatorname{Re}(E(r,p)-g)_m] \operatorname{Re}[\partial_{p_n} (E(r,p))_m] \\
&\quad + [\operatorname{Im}(E(r,p)-g)_m] \operatorname{Im}[\partial_{p_n} (E(r,p))_m]\} && (9.35) \\
&\overset{(9.28)}{=} \frac{1}{\sigma^2} \sum_{m=1}^{M} \{[\operatorname{Re}(E(r,p)-g)_m] \operatorname{Im}[\overline{r_n B_{m,n} \exp(\mathtt{i} p_n)}] \\
&\quad + [\operatorname{Im}(E(r,p)-g)_m] \operatorname{Re}[\overline{r_n B_{m,n} \exp(\mathtt{i} p_n)}]\} && (9.36) \\
&= \operatorname{Im}\{(BR\Phi)^* [E(r,p)-g]\}_n = \operatorname{Im}\{R\Phi^{-1} B^* [E(r,p)-g]\}_n. && (9.37)
\end{aligned}$$

■

9.2.4. Analysis of the Non-Linear Data Term. We will know take a look at the non-linear data term.

Lemma 9.2.3. *Let* $\Phi := \operatorname{diag}(\exp(\mathtt{i} p)), R := \operatorname{diag}(r), \tilde{I} = B^* B$ *and* $D := \operatorname{diag}(\Phi^{-1} \tilde{I} \Phi r - \Phi^{-1} B^* g)$. *Then the Hessian of* $\mathcal{E} \colon \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ *defined as* (9.18)

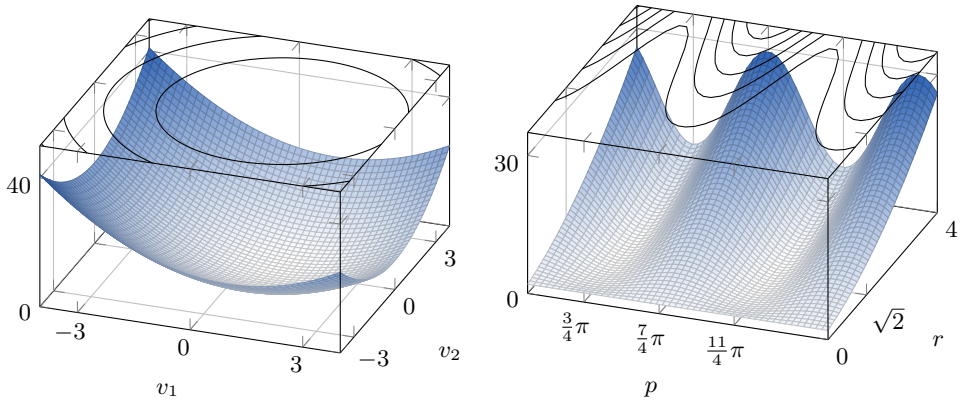


Figure 9.18. *The MRI fidelities for linear parametrization* (9.7) *on the left and non-linear parametrization in terms of magnitude and phase* (9.18) *on the right. The data has been chosen to be* $g = \sqrt{2}\exp(\frac{3}{4}\pi\mathtt{i})$.

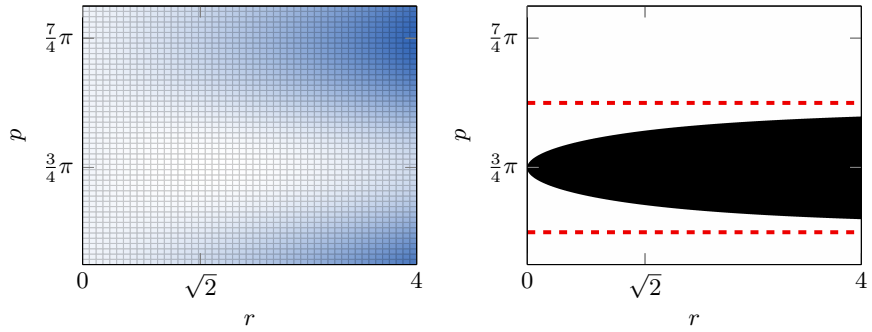


Figure 9.19. *The data fidelity with magnitude and phase parametrization* (9.18) *is globally non-convex but convex on the set* U *shown on the right. It is visualized for* $g = \sqrt{2}\exp(\frac{3}{4}\pi\mathtt{i})$ *and the dashed lines are the bounds* $\frac{3}{4}\pi \pm \frac{1}{2}\pi$.

is given by

$$H\mathcal{E}(E(r,p)|g) = \frac{1}{\sigma^2}\begin{pmatrix} \operatorname{Re}(\Phi^{-1}\tilde{I}\Phi) & \operatorname{Im}(R\Phi^{-1}\tilde{I}\Phi + D) \\ \operatorname{Im}(R\Phi^{-1}\tilde{I}\Phi + D) & \operatorname{Re}(R\Phi^{-1}\tilde{I}\Phi R - RD) \end{pmatrix}. \tag{9.38}$$

Proof. The first derivatives are obtained from lemma 9.2.2 as

$$\partial_r \mathcal{E}(E(r,p)|g) = \frac{1}{\sigma^2}\operatorname{Re}\{\Phi^{-1}B^*[E(r,p) - g]\} \tag{9.39}$$

$$\partial_p \mathcal{E}(E(r,p)|g) = \frac{1}{\sigma^2} R \operatorname{Im}\{\Phi^{-1}B^*[E(r,p) - g]\}. \tag{9.40}$$

For the second derivatives, we get

$$\sigma^2 \partial_{r_l,r_k}\mathcal{E}(E(r,p)|g) = \partial_{r_l}\operatorname{Re}\{\Phi^{-1}B^*[E(r,p) - g]\}_k \tag{9.41}$$

$$= \operatorname{Re}\{\partial_{r_l}(\Phi^{-1}B^*B\Phi r - \Phi^{-1}B^*g)_k\} = \operatorname{Re}(\Phi^{-1}\tilde{I}\Phi)_{k,l} \tag{9.42}$$

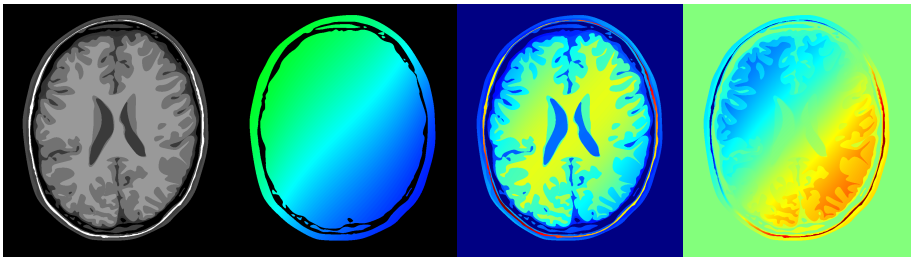


Figure 9.20. *From left to right: magnitude, phase, real part and imaginary part of a simulated MRI image which serves as a ground truth in our experiments.*

$$\sigma^2 \partial_{r_l,p_k} \mathcal{E}(E(r,p)|g) \tag{9.43}$$

$$= \partial_{r_l} \Big\{ R \operatorname{Im}\{\Phi^{-1} B^* [E(r,p) - g]\} \Big\}_k \tag{9.44}$$

$$= \operatorname{Im}\{\Phi^{-1} B^* [E(r,p) - g]\}_k \delta_{k,l} + r_k \operatorname{Im}\{\partial_{r_l} (\Phi^{-1} B^* B \Phi r - \Phi^{-1} B^* g)_k\} \tag{9.45}$$

$$= \operatorname{Im}[D_{k,l} + (R\Phi^{-1} \tilde{I} \Phi)_{k,l}] \tag{9.46}$$

$$\sigma^2 \partial_{p_l,p_k} \mathcal{E}(E(r,p)|g)$$

$$= \partial_{p_l} \Big\{ R \operatorname{Im}\{\Phi^{-1} B^* [E(r,p) - g]\} \Big\}_k \tag{9.47}$$

$$= r_k \operatorname{Im}\Big\{ \partial_{p_l} \{\Phi^{-1} B^* [BR \exp(\mathtt{i}p) - g]\}_k \Big\} \tag{9.48}$$

$$= r_k \operatorname{Im}\Big\{ -\mathtt{i}\{\Phi^{-1} B^* [E(r,p) - g]\}_k \delta_{k,l} + \mathtt{i}(\Phi^{-1} B^* BR\Phi)_{k,l} \Big\} \tag{9.49}$$

$$= \operatorname{Re}(R\Phi^{-1} \tilde{I} \Phi R - RD)_{k,l}. \tag{9.50}$$

■

Special case: B is unitary. In the case when B is unitary, i.e. $B^*B = I$, then there is $\tilde{I} = I$ and the Hessian simplifies significantly. We can show that although the data fidelity is non-convex globally, it is convex indeed on a relatively large set.

Proposition 9.2.4 (Non-Convexity). *Let $g \neq 0$. Then the function $\mathcal{E}$ defined by* (9.18) *is not convex.*

Proof. Let $g = z \exp(\mathtt{i}\theta) \neq 0$ with $z \neq 0$. It follows from lemma 9.2.3 that Hessian with $\tilde{I} = I$ is given by

$$H\mathcal{E}(E(r,p)|g) = \frac{1}{\sigma^2} \begin{pmatrix} I & \operatorname{diag}[z \sin(p-\theta)] \\ \operatorname{diag}[z \sin(p-\theta)] & \operatorname{diag}[rz \cos(p-\theta)] \end{pmatrix}. \tag{9.51}$$

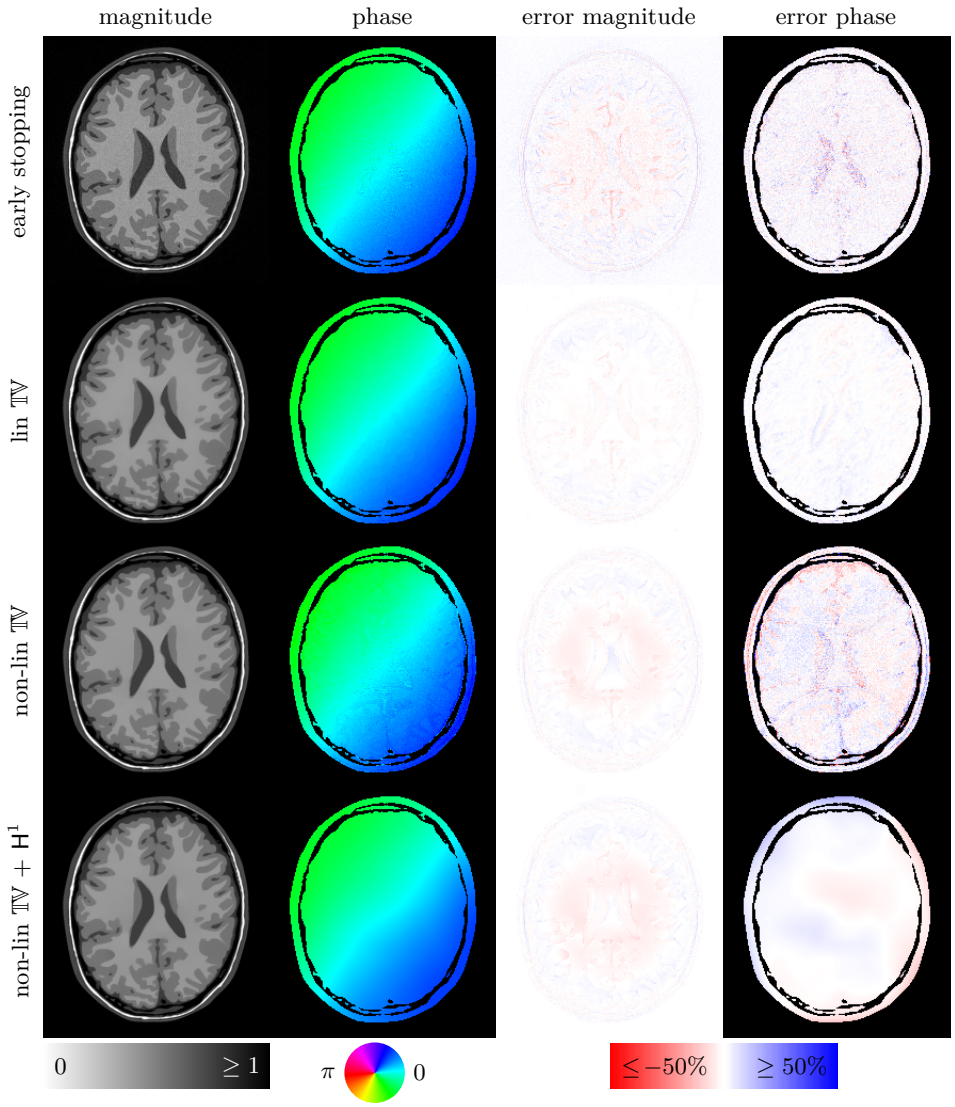


Figure 9.21. *Results for* `lines2`*, i.e. sampling every 2nd line in k-space. Parameters are optimized for smallest* ℓ^2*-error of the magnitude.*

Then for any $x, y \in \mathbb{R}^N$, there is

$$\begin{aligned}&[x;y]^T H\mathcal{E}(E(r,p)|g)[x;y]\\ &\quad = |x|^2 + \sum_{n=1}^{N} \{y_n^2 r_n z_n \cos(p_n - \theta_n) + 2x_n y_n z_n \sin(p_n - \theta_n)\}. \end{aligned} \tag{9.52}$$

Let $\hat{p} = \theta + \pi$ and $\hat{r} \equiv 1$. Then, with $x = 0$, $y = \sqrt{z}$ there is

$$[x;y]^T H\mathcal{E}(E(r,p)|g)[x;y] = -|z|^2 < 0 \tag{9.53}$$

as $g \neq 0$, so that $\mathcal{E}$ is not convex. ∎

Proposition 9.2.5 (Local Convexity). *There is a non-trivial convex set* $\mathsf{U} \subset [0,\infty)^N \times$

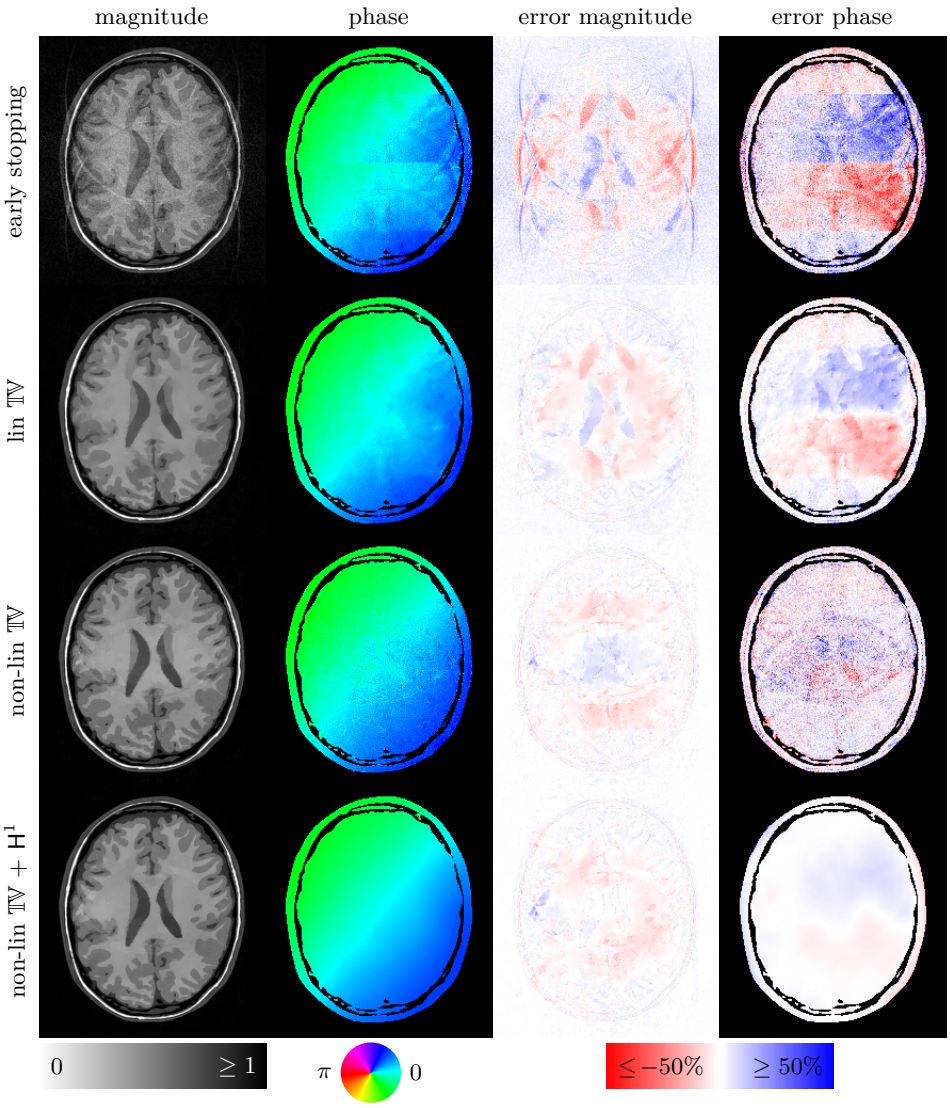


Figure 9.22. *Results for* `lines4`*, i.e. sampling every 4th line in k-space. Parameters are optimized for smallest* ℓ^2*-error of the magnitude.*

$[0, 2\pi)^N$ *so that* $\mathcal{E}$ *defined by* (9.18) *restricted to* U *is convex.*

Proof. Let $g = z \exp(\mathrm{i}\theta) \in \mathbb{C}^N$ and denote the periodic ball of radius R around $y \in \mathbb{R}$ by $\mathbb{B}_R(y) := \{x \in \mathbb{R} \mid \min_{k\in\mathbb{Z}} |x - y + 2\pi k| < R\}$. Furthermore, let

$$\mathsf{U} := \Big\{(r, p) \in [0, \infty)^N \times [0, 2\pi)^N \mid r_n \geq x_n \Delta_n(p_n) \text{ and } p_n \in \mathbb{B}_{\pi/2}(\theta_n)\Big\} \tag{9.54}$$

with $\Delta_n(x) := \sin^2(x - \theta_n)/\cos(x - \theta_n) > 0$. Straightforward computations show that

$$\Delta_n''(x) = 2\cos(x - \theta_n) + \Delta_n(x) + 2\tan^2(x - \theta_n)/\cos(x - \theta_n) > 0 \tag{9.55}$$

for $p_n \in \mathbb{B}_{\pi/2}(\theta_n)$, hence, Δ_n is convex and so is U.

It remains to show that $\mathcal{E}|_{\mathsf{U}}$ is convex. For any $(r, p) \in \mathsf{U}$ and $x, y \in \mathbb{R}^N$ there is

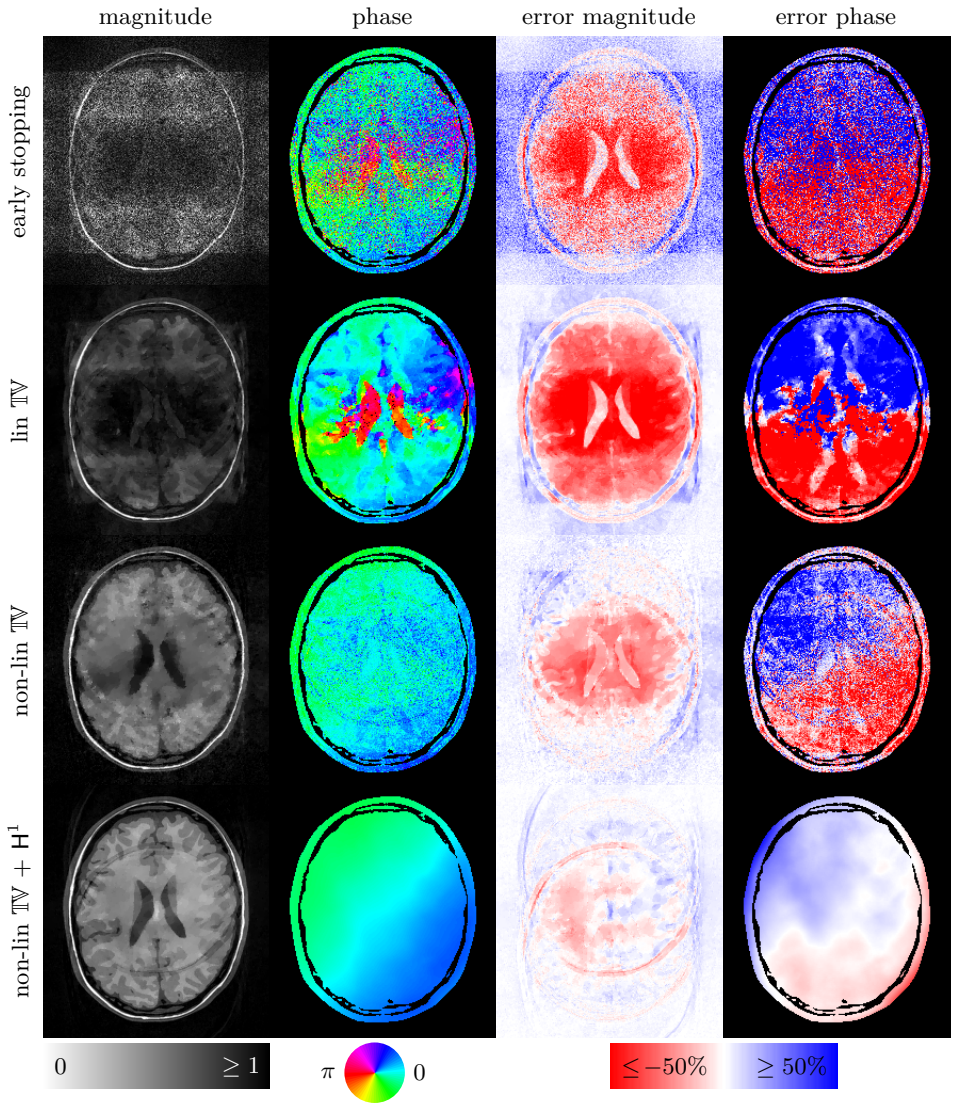


Figure 9.23. *Results for* `lines6`*, i.e. sampling every 6th line in k-space. Parameters are optimized for smallest ℓ^2-error of the magnitude.*

with (9.52)

$$(x;y)^T H\mathcal{E}(E(r,p)|g)(x;y) \geq \sum_{n=1}^{N} y_n^2 z_n r_n \cos(p_n-\theta_n) - y_n^2 z_n^2 \sin^2(p_n-\theta_n) \tag{9.56}$$

$$\geq \sum_{n=1}^{N} y_n^2 z_n \cos(p_n-\theta_n)[r_n - z_n \Delta_n(p_n)] \geq 0. \tag{9.57}$$

This means that the Hessian is positive-semidefinite and hence, $\mathcal{E}|_{\mathsf{U}}$ is convex. ∎

Remark 9.2.6. While the first proposition tells us that the minimization problem has become much harder, the second proposition ensures that as long as we are not too far away from the data (namely being in the set U shown in figure 9.19 on the right) we are guaranteed to converge to a global solution using a gradient type optimization

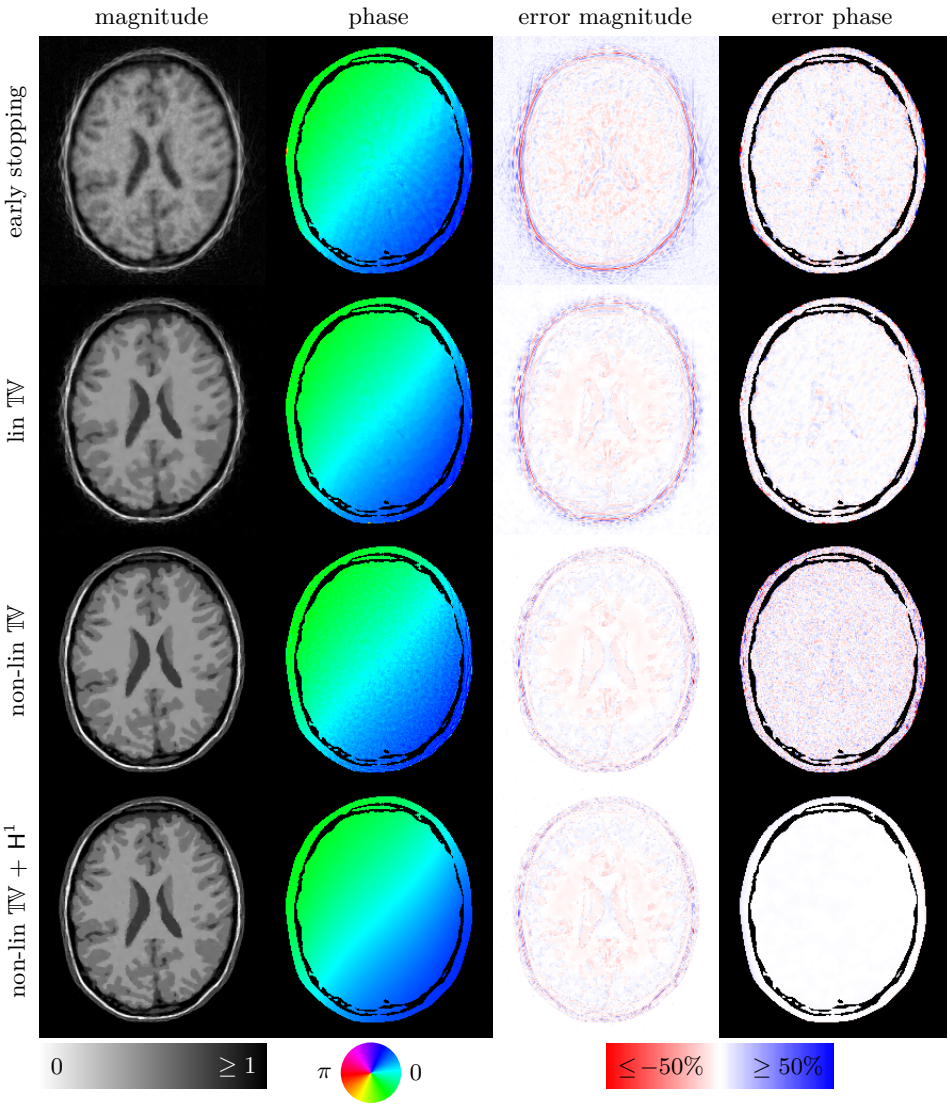


Figure 9.24. *Results for* `radial20`*, i.e. sampling along 20 radial spokes. Parameters are optimized for smallest* ℓ^2*-error of the magnitude.*

scheme such as gradient descent. ▲

Remark 9.2.7. The results have been proven only for unitary operators which corresponds to MRI reconstruction with complete data. In the extreme case when no data is acquired, the operator becomes constant zero and the data fidelity is trivially convex everywhere. Therefore, we anticipate that for every kind of undersampling in MRI, the non-linear data fidelity is convex on a non-trivial subset. ▲

9.2.5. Numerical Simulation.

Setting. We now present numerical results for MRI reconstruction from undersampled data when the ground truth is complex, see figure 9.20. We compare four different

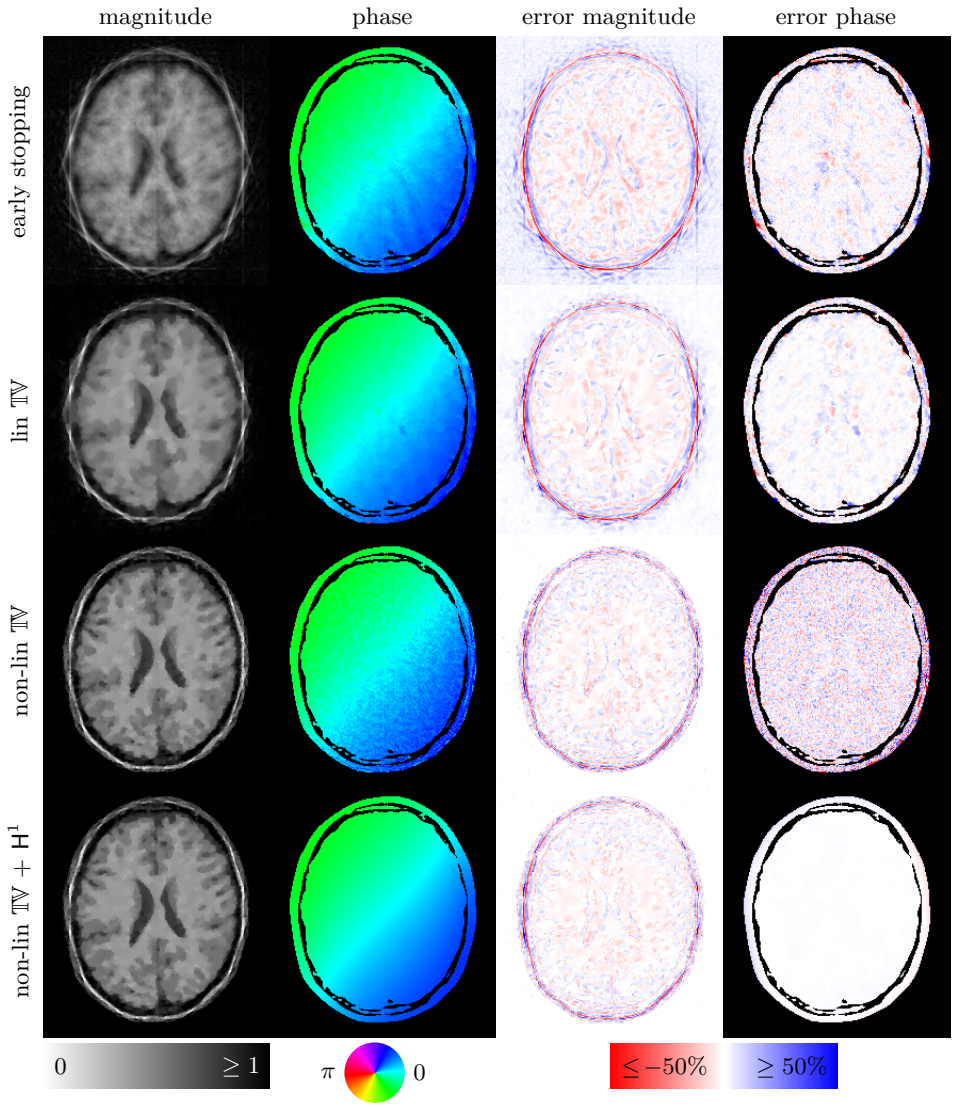


Figure 9.25. *Results for* `radial10` *i.e. sampling along 10 radial spokes. Parameters are optimized for smallest ℓ^2-error of the magnitude.*

methods: i) linear parametrization with early stopping, ii) linear parametrization with total variation on the complex image as a prior, iii) non-linear parametrization with total variation on the magnitude and the phase is not regularized and iv) non-linear parametrization with total variation on the magnitude and the phase regularized with the H^1-seminorm. Note, that the phase is modelled to be smoothly varying in $[-1, 1]$ and there is no phase wrapping.

Data. The simulated magnitude is based on a BrainWeb [40] phantom that has been transformed into a continuous phantom [71]. We sample from this continuous phantom at a very fine grid of size 1024×1024. The phase is simulated to be smoothly varying with values between -1 and 1. It is desirable to extend this setting to faster

varying phase with a lot larger values such that phase wraps occur. The simulated magnitude and phase are shown in figure 9.20. While the data was generated with a high-resolution phantom, we reconstruct on a 256×256 grid to avoid an inverse crime. The data is given for eight coils with sensitivities computed by the software from [71].

Optimization. In all four cases, optimizations are performed with the bounded large-scale quasi-Newton method L-BFGS-B [24]. When the images are parametrized with magnitude and phase we bound the magnitude to be non-negative.

Results. The results are shown in figures 9.21-9.25. As it can be seen, all four methods do well if we sample the k-space at every 2nd line. For larger Cartesian undersampling we see that penalizing the magnitude with total variation is far superior than applying total variation on the complex image. Especially for sampling every 6th line, getting a better phase estimate by using the H^1-seminorm also yields better reconstructions of the magnitude.

For radial sampling, similar observations can be made although this time the difference between the methods is far less severe.

9.2.6. Conclusions of this section. The prior knowledge for sparse MRI is usually stated on the complex image. We reformulate the problem so that the prior knowledge can be stated directly on the magnitude and phase. We showed that the new parametrization – proven only for unitary operators – leads to a locally convex problem and is therefore well-behaved in a local region. The numerical results clearly indicate that penalizing the magnitude and phase rather than the complex image itself leads to less artefacts.

9.3. Joint Reconstruction for PET-MRI with Phase and Parallel MRI.

In this section we present results for joint reconstruction for PET-MRI with a more complicated model for MRI. First, the ground truth in MRI is modelled to have both magnitude and phase. The two modalities are coupled within the joint reconstruction by assuming structural similarity between the activity image in PET and the magnitude in MRI. Second, several receiver coils are employed to acquire the data for MRI in parallel. This work has been published in [52].

9.3.1. Materials and Methods.

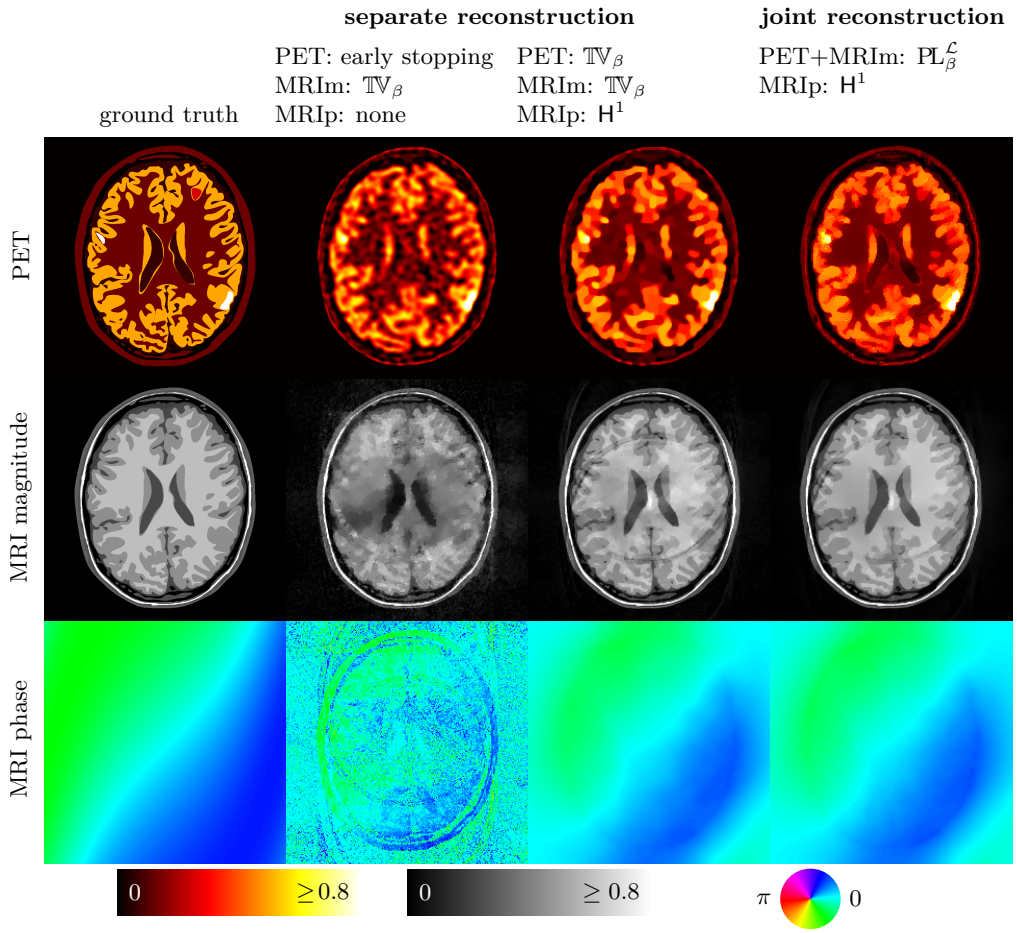


Figure 9.26. *PET and MRI (both magnitude and phase) reconstruction from Cartesian sampling where only every 6th line is acquired. Coupling PET activity with MRI magnitude with the parallel level sets prior gives less under-sampling artefacts in MRI and a better defined PET image.*

Methods. We will compare separate regularization to joint reconstruction. For separate PET reconstruction, we either perform early stopping or use a total variation prior. In the case of separate MRI reconstruction, we always have a total variation prior on the magnitude but once with and once without H^1-regularization on the phase. For joint reconstruction, we consider the case that the PET activity is coupled to the MRI magnitude via the linear parallel level sets prior and the phase is H^1-regularized.

Data. The phantom is similar to the phantom in section 9.1 but the lesions are placed at different locations. We reconstruct images of 256 $\times$ 256 pixels with a pixel size of 1mm $\times$ 1mm. The PET resolution is modelled by a Gaussian blur with FWHM of 5mm $\times$ 5mm. We modelled attenuation but no scatter or randoms. We considered two cases of under-sampling in MRI. In the first case, we omitted five out of six phase encodings in the Cartesian sampling which is a speed up of more than 80%. In the second case the data were sampled only at ten radial spokes.

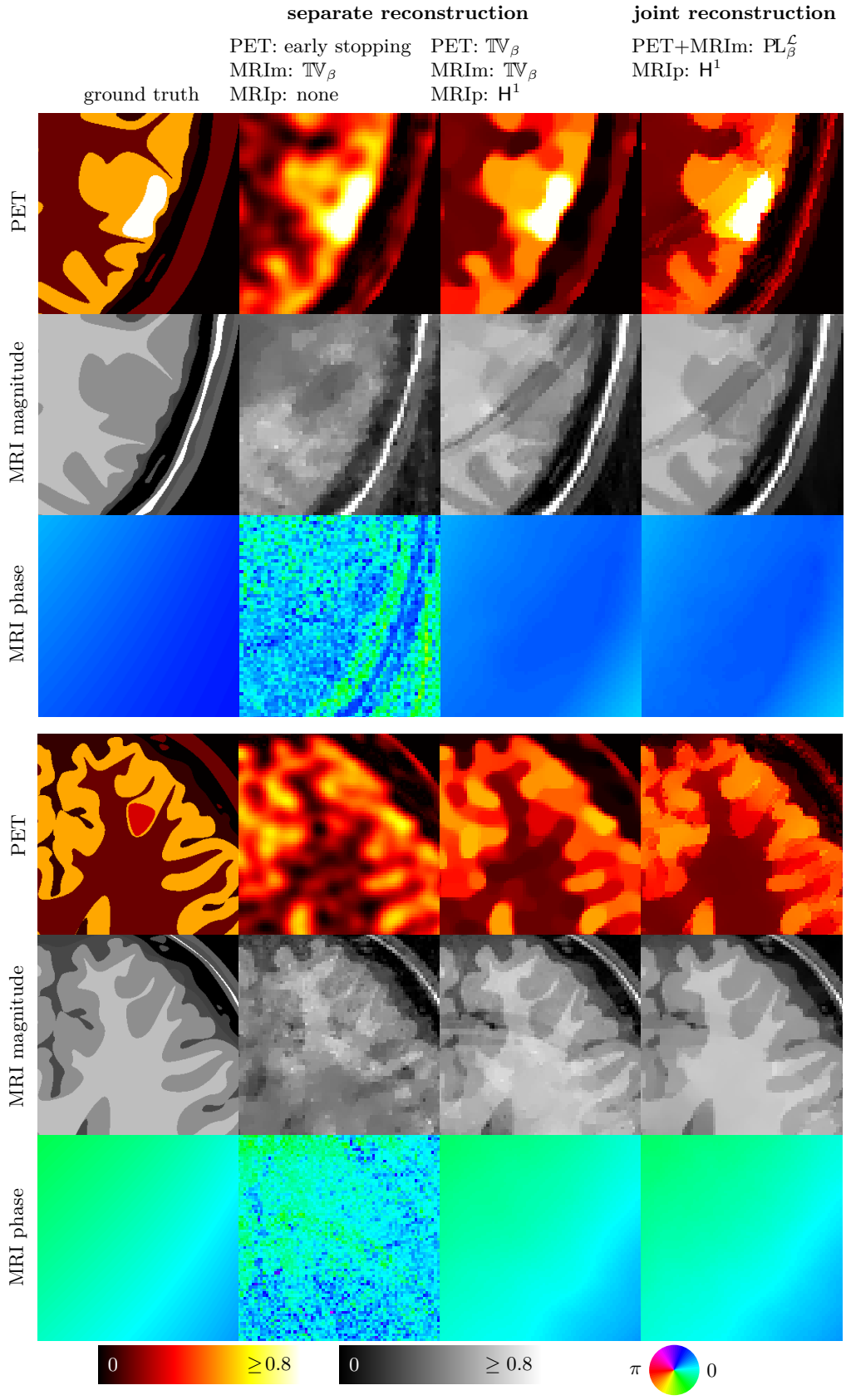


Figure 9.27. *Close-ups of figure 9.26 on the cold lesion (top) and right hot lesion (bottom).*

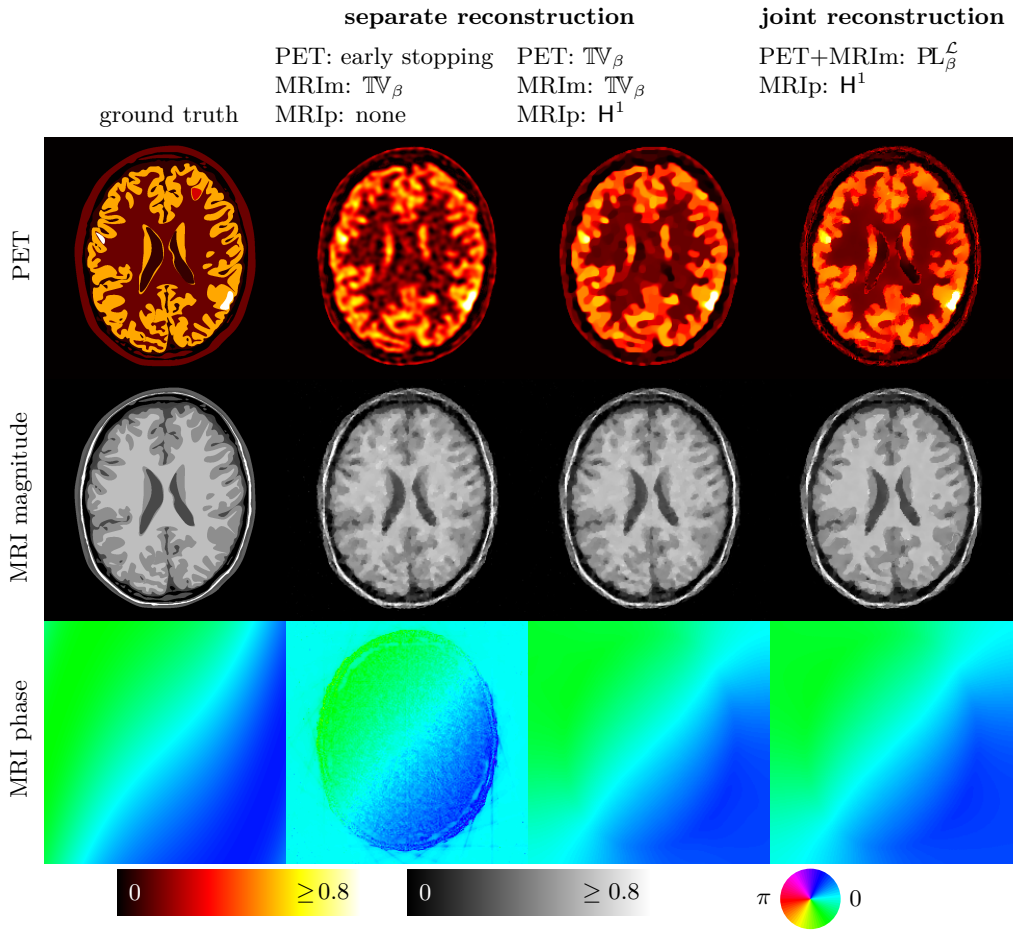


Figure 9.28. *PET-MRI reconstruction where the MRI is sampled at 10 radial spokes. The prior on the phase allows a good reconstruction of the phase itself but does not seem to have much impact on the magnitude (compare second and third column). By combining the PET activity with the MRI magnitude much better defined images for both modalities can be obtained.*

Optimization. As in previous examples, we employ L-BFGS-B for the minimization of the functionals [24, 117, 113].

Parameter Choice. The two parameters for regularization α and smoothing β are varied over a range of suitable values. We will pick the parameter constellation that results in the smallest ℓ^2-error between the estimated images and the ground truth.

9.3.2. Results. Figure 9.26 shows the results for Cartesian sampling. It can be seen from the two middle columns that penalizing the phase of an MRI image has a huge effect on the magnitude as well. Furthermore, using total variation in the PET reconstruction yields better defined regions. Joint reconstruction of the MRI magnitude and the PET activity with the parallel level sets prior results in a much better defined PET image. In addition some under-sampling artefacts in the MRI magnitude can be reduced. This can also be seen in the close-ups in figure 9.27.

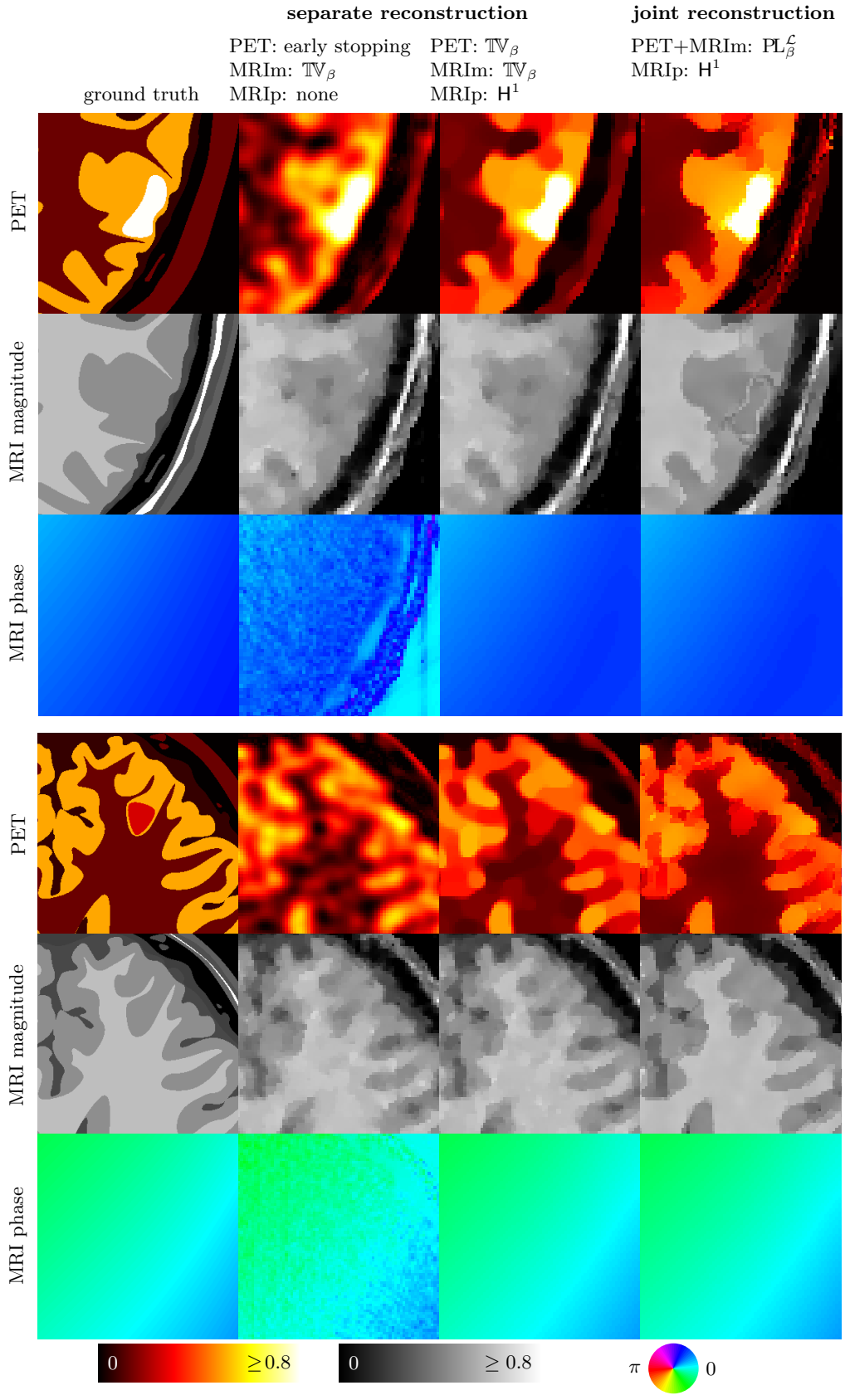


Figure 9.29. *Close-ups of figure 9.28 on the cold lesion (top) and right hot lesion (bottom).*

The reconstructed images for the radial sampling can be seen in figure 9.28 with close-ups in figure 9.29. Although the undersampling artefacts for radial sampling are very different qualitatively similar results are obtained.

9.4. Summary.

In this chapter we presented results where we jointly reconstructed PET and MRI. We have seen that coupling these two modalities leads to better defined PET images. Moreover, in many situation the coupling was also helpful to reduce undersampling artefacts in MRI. In addition, we presented a modified parametrization for MRI in terms of magnitude and phase which is useful as our prior knowledge on MRI is naturally on these parameters. With this parametrization we can link the PET activity to the magnitude in MRI and therefore jointly reconstruct PET-MRI even when phase is present.

Chapter 10

Discussion, Conclusions and Open Problems

10.1. Summary.

In this thesis, we contributed to the field of multi-modality imaging. We developed a framework that can exploit the intrinsic common structure often present in multi-modality imaging applications. In this framework, we performed joint reconstruction as we combined the modalities by joint prior information in a joint minimization problem. We presented a graphical model based on intuitive beliefs that yields the functional that has been used for joint reconstruction by other authors. The joint prior knowledge is based on the expected common structure in the images of the different modalities. We proposed a family of symmetric and asymmetric parallel level set priors that can measure common structure and therefore can be used to promote similar structures in images of different modalities.

We have analysed these families with respect to their corresponding diffusive flows. In contrast to most priors, the diffusive flow for both symmetric and asymmetric parallel level sets is anisotropic as it depends on the local coordinates of the image of other modality. Moreover, we analysed the parallel level set families in terms of convexity. It turns out that none of these priors is convex in the joint argument as they all fail the necessary growth condition. In case of quadratic parallel level sets, the functional is convex in the first / each argument for the asymmetric / symmetric case.

We have shown in numerous examples that joint reconstruction can be beneficial compared to separate reconstruction. First, by coupling the colour channels in RGB colour imaging, the results for both denoising and demosaicking can be significantly improved. This is especially relevant to avoid colour smearing which results otherwise in artefactual colour artefacts. Second, we have performed PET-MRI reconstructions

with highly undersampled MRI data. As many other methods, the proposed framework is capable of transferring the high spatial resolution from MRI to PET such that the reconstructed PET images have clearly defined details. Moreover, the MRI images obtained from joint PET-MRI reconstruction often have a significant reduction of undersampling artefacts. These results are consistent under several software phantoms and MRI data samplings. Similar results were obtained for a realistic scenario where we considered parallel MRI and included phase in MRI.

These results have been published in journals [50, 51] and in conference proceedings [52]. Moreover, they have been presented at numerous national and international conferences.

10.2. Discussion.

The premise of this thesis is that there is structural coupling of the images of some modalities that we can exploit. This is especially helpful if no physical model is known that relates these images. In order for joint reconstruction to be beneficial, there needs to be a significant amount of common structure. There is not much doubt that this assumption holds for most natural images in colour photography. However, the structural similarity for medical images – especially between structure and function – might not always been present. In particular in PET, the reconstructed image is highly dependent on the tracer, so that for some tracers the anatomy shows up and the structural similarity with MRI holds while for other tracers this might not be the case. Likewise, MRI images can be corrupted by geometric distortions, artefacts from susceptibility and eddy currents. It is important to correct for such artefacts when joint reconstruction shall be employed.

If the premise of structural coupling holds, then the framework in this thesis is applicable and we can exploit the common structure within a joint reconstruction. The joint prior knowledge is based on the expected common structure in the images of the different modalities. To model the structure of an image by means of its level sets and therefore to measure common structure with parallel level sets seems to be a good approach. It is a simple model that fulfils the requirement to be "intensity" independent as it is not the actual "colours" that matter. We proposed a family of symmetric parallel level set priors that can measure the common structure defined in this sense. Although its extension as a regularization with the smoothing parameter β for the smoothed norms is somewhat ad hoc, it gives reasonable results. Both the

theoretical observations in chapter 5 as well as our numerical examples in chapter 9 show that linear parallel level sets does not only promote joint structure but even enforces it. While it results in visually appealing results for RGB colour imaging, it might not be reasonable for medical images that will eventually be used in clinical diagnostics. In addition, all the discussed models are non-convex in the joint argument. While for quadratic parallel level sets this does not seem to matter as the same numerical results were obtained from very different initial conditions, this was unfortunately the case for linear parallel level sets. Again, in RGB colour imaging this is not very problematic as the initial data is already close to the solution we seek.

The joint reconstruction framework with symmetric parallel level sets is intrinsically symmetric. Not only is the prior symmetric in its arguments but so is the regularization parameter that is the same for both modalities. This is conceptually appealing and has the advantage that only a single parameter needs to be determined. However, in some cases this could lead to suboptimal results. If one wants to practically combine the reconstruction of two modalities, it needs to be better than both of them individually and it might be that such a simple model can not live up to such expectations.

As it is the case for all variational models, the regularization parameter plays an important role and its selections is not always easy without extra knowledge about the problem. Nevertheless, if the imaging set-up is similar one can expect that a similar regularization parameter serves as a good candidate such that its order of magnitude can be known a priori for a fixed experimental set-up. In addition to the overall regularization parameter, the parallel level set priors have another parameter that controls the smoothness of the functional and adds regularization to it. Similar to the smoothing parameter we employed for smooth total variation, the parameter should be selected proportional to the gradient magnitudes we expect in the reconstructed image. For the examples we considered, good choices were in the order of 0.1% - 1% of the largest expected gradient magnitudes.

10.3. Future Work.

In this section we will discuss possible future lines of research that directly extend the work presented in this thesis.

Although the framework that we presented is quite general, our results concentrated on using parallel level sets to encode structure in the image. Other priors could be tested as well. For instance total nuclear variation – Schatten 1-norm – we briefly discussed

in section 5.5, This prior satisfies our general requirement for encoding structure while also reducing to total variation. In addition to these properties it is – as a seminorm on the joint image – convex in the joint argument. It has been used for RGB imaging [80] and multi-energy CT reconstructions [133]. To date there has not been a comparison between our parallel level set model and total nuclear variation. It is also not clear whether this can couple the modalities strongly enough.

Another possible direction is to perform joint reconstruction with explicit segmentation. The segmentation should be based on both images and could either be posed as a hierarchical model [125, 123] or based on level sets [42, 38]. It has already been used for one-side reconstruction [100] but to the best of our knowledge there was no attempt to extend this approach to joint reconstruction. The asymmetric parallel level sets model can for instance be used to reconstruct an image when an explicit segmentation is given.

In the proposed joint model, there is one prior term that couples regularization and structural similarity. Another option is to decouple these and have multiple priors that incorporate different a priori information as in [83, 102, 155]. This approach has been used as well for joint reconstruction of PET-MRI [94].

In the numerical examples for RGB demosaicking and PET-MRI with undersampled MRI we have used different sampling patterns. A sampling pattern for RGB demosaicking is the Bayer filter, cf. figure 8.1, but other patterns are also possible. For MRI reconstruction we have chosen different sampling patterns like line-by-line Cartesian sampling, radial sampling or spiral sampling. In all of these cases, there is a large degree of freedom on how we sample. Not only can we choose any of these schemes but each of them has also several parameters that can be tuned. An “optimal” sampling pattern will depend on the image content itself and on the prior knowledge we exploit. Therefore, it would be interesting to find an “optimal” sampling pattern that allows good reconstruction for a large class of images given a certain a priori knowledge. Note, that this is related to optimal experimental design, e.g. [77], and the bi-level approach to image compression [36].

As we discussed before, some of the optimization problems are non-convex (due to the non-convex prior). Nevertheless, in this thesis we tried to solve these with gradient based optimization. Thus, we can only expect local optimality in the sense that we have a stationary point, i.e. the gradient vanishes. Probably, it would be beneficial to use multi-resolution or continuation techniques similar to image registration.

The numerical results showed that the linear parallel level set prior performed extremely well in regions of common structure. In regions with different structure the prior either smoothed out the unique information of one modality or created a false edge in the other. An option to circumvent this would be a local regularization parameter choice. Based on an estimate whether we are in a region of common structure or not, the regularization parameter could be chosen adaptively. This has been used already for total variation reconstruction where the regularization had to be chosen adaptively in order to preserver fine details, cf. [78, 95, 45, 21] and references therein.

The most common approach to joint reconstruction is to minimize a joint objective functional as we did in this thesis. We have already mentioned in the discussion section that this symmetric approach might be suboptimal if the pair of modalities is very asymmetric. On the one hand, every modality might need their own regularization parameter. On the other hand, for every modality it might be different what should happen in the degenerate case when no side information is present. The family of asymmetric parallel level sets is flexible enough to allow for all these peculiarities in a one-sided reconstruction. However, it is not yet clear how it can be used for joint reconstruction.

The numerical results in this thesis were all simulations and as initial findings limited to a rather simple setting. Therefore, it would be very interesting to study the behaviour of joint reconstruction with respect to noise: In the case of PET for different number of observed counts and in the case of MRI for an increasing standard deviation of the Gaussian noise. Moreover, results from hardware phantoms as well as clinical data will shed light on the applicability of joint reconstruction in a practical setting and how robust this approach is with respect to imperfections.

In this thesis we presented results for the multi-modality imaging examples of RGB imaging and PET-MRI. However, the framework is independent of the actual modalities and thus can be used for other combinations of modalities. A promising example for joint reconstruction is for instance multi-contrast MRI reconstruction [13]. As the resolution is the same for any MRI contrast, it is intuitive that we can infer structural information in both directions. In addition, we have seen in the beginning of this thesis that MRI is an active imaging device: we can choose where in the k-space we sample. Therefore, it might be possible to have two complementary sampling pattern that avoid redundant data.

10.4. Conclusions.

Overall, the general framework, the methods and the numerical results in this thesis are a big leap forward for joint reconstruction in multi-modality imaging. This generic set-up can be used for any combination of modalities and is not limited to colour or medical imaging. The family of parallel level sets priors can link different modalities by structure and therefore empowers us to exploit the intrinsic common structure seen in many multi-modality imaging applications. Furthermore, the numerical results showed that this framework can indeed utilize similar structures in multi-modality imaging so that joint reconstruction can be beneficial to all modalities. Thus, we showed that "one plus one can add up to more than two" [39].

List of Symbols

Numbers

$\mathbb{N}$	natural numbers $1, 2, \dots$
$\mathbb{Z}$	integers, $\dots, -2, -1, 0, 1, 2, \dots$
$\mathbb{R}$	real numbers
$\mathbb{C}, \mathrm{Re}, \mathrm{Im}, \overline{x}$	For a complex number $z = x + \mathtt{i}y \in \mathbb{C}$ with $x, y \in \mathbb{R}$ and the imaginary unit, i.e. $\mathtt{i} \coloneqq \sqrt{-1}$, the real and imaginary parts are $x = \mathrm{Re}(z), y = \mathrm{Im}(z)$ and its complex conjugate is $\overline{z} \coloneqq \mathrm{Re}(x) - \mathtt{i}\,\mathrm{Im}(x)$.

Functions and Spaces

$\mathsf{C}^p(\mathsf{X}, \mathsf{Y}), \mathsf{C}(\mathsf{X}, \mathsf{Y})$	space of p times continuously differentiable mappings from X to Y, $\mathsf{C} \coloneqq \mathsf{C}^0$, $\mathsf{C}^p(\mathsf{X}) \coloneqq \mathsf{C}^p(\mathsf{X}, \mathbb{R})$
$\Omega, \lvert\Omega\rvert$	image domain $\Omega \subset \mathbb{R}^D$ and its "size" $\lvert\Omega\rvert \coloneqq \int_\Omega 1$
$\partial\Omega, \nu$	boundary of domain Ω and outer normal vector ν
$\mathsf{C}^p_*(\Omega, \mathsf{Y})$	$\mathsf{C}^p(\Omega, \mathsf{Y})$ with Neumann boundary conditions, i.e. if $u \in \mathsf{C}^p_*(\Omega, \mathsf{Y})$, then $\partial_\nu u\vert_{\partial\Omega} = 0$
$\mathsf{H}^{p,m}(\mathsf{X}, \mathsf{Y}), \mathsf{H}^p(\mathsf{X}, \mathsf{Y})$	Sobolev space of p times weakly differentiable and m times integrable mappings from X to Y, $\mathsf{H}^p(\mathsf{X}, \mathsf{Y}) \coloneqq \mathsf{H}^{p,2}(\mathsf{X}, \mathsf{Y})$, $\mathsf{H}^{p,m}(\mathsf{X}) \coloneqq \mathsf{H}^{p,m}(\mathsf{X}, \mathbb{R})$
$\mathsf{L}(\mathsf{X}, \mathsf{Y})$	space of linear and continuous mappings from X to Y
A^*	adjoint of the linear operator/ matrix A
id, I	the identity function and identity matrix / operator

Derivatives

u', u''	1st and 2nd derivative of a function $u\colon \mathbb{R} \to \mathbb{R}$
$\partial_x u, \partial_n u$	partial derivative of a function u with respect to x and the nth component
$\partial_{n,m} u$	mixed partial derivatives, $\partial_{n,m} u \coloneqq \partial_n \partial_m u$
∇	gradient of a function $u\colon \mathbb{R}^N \to \mathbb{R}$, $(\nabla u)_n(x) \coloneqq \partial_n u(x)$

Dw	Jacobian of a function $w\colon \mathbb{R}^N \to \mathbb{R}^M$, $Dw \coloneqq [\nabla w_1, \dots, \nabla w_M]$
$\operatorname{div} w$	divergence of a function $w\colon \mathbb{R}^N \to \mathbb{R}^N$, $\operatorname{div} w(x) \coloneqq \sum_{n=1}^N \partial_n w_n(x)$
Hu	Hessian of a function $u\colon \mathbb{R}^N \to \mathbb{R}$, $(Hu(x))_{n,m} \coloneqq \partial_{n,m} u(x)$
$D\mathcal{P}[u]$	Gâteaux derivative of $\mathcal{P}$ at u
$\partial^c u$	central differences on the usual grid $\partial^c u_n \coloneqq 1/2(u_{n+1} - u_{n-1})$ or on a subgrid $\partial^c u_n \coloneqq u_{n+1/2} - u_{n-1/2}$
$\partial^+ u, \partial^- u$	forward and backward differences, $\partial^+ u_n \coloneqq u_{n+1} - u_n$, $\partial^- u_n \coloneqq u_n - u_{n-1}$
$\nabla^+ u, \nabla^c u$	gradient approximation with forward / central differences
$\operatorname{div}^- u, \operatorname{div}^c u$	divergence approximation with backward / central differences

Specific Regularizations

$\lvert\cdot\rvert_{\mathsf{H}^1}$	seminorm of $\mathsf{H}^1(\Omega)$
$\mathbb{TV}, \mathbb{TV}^{\mathcal{J}}, \mathbb{TV}^{\mathcal{C}}, \mathbb{TV}^{\mathcal{S}}$	usual, joint, colour and separate total variation
$\mathbb{TV}_\beta, \mathbb{TV}_\beta^{\mathcal{J}}, \dots$	smoothed total variation, etc
$\mathrm{PL}, \mathrm{PL}^{\mathcal{L}}, \mathrm{PL}^{\mathcal{Q}}$	general, linear, quadratic (symmetric) parallel level sets
$\mathrm{APL}, \mathrm{APL}^{\mathcal{L}}, \mathrm{APL}^{\mathcal{Q}}$	general, linear, quadratic asymmetric parallel level sets
$\mathrm{PL}_\beta^{\mathcal{L}}, \mathrm{PL}_\beta^{\mathcal{Q}}, \mathrm{APL}_\beta^{\mathcal{L}}, \cdots$	smoothed linear parallel level sets, etc
CG	cross-gradients functional
$f_{\mathrm{PL}}, f_{\mathrm{PL}^{\mathcal{L}}}, \cdots$	local regularization functions

Diffusive Flow

$\kappa, \varkappa$	isotropic and anisotropic diffusivities
$\operatorname{span}(x), \operatorname{orth}(x)$	span and orthogonal complement of a vector x
$\lambda^{\parallel}, \lambda^{\perp}$	Eigen values of the diffusivity parallel and orthogonal to the level sets of the underlying structure
R_u	local (Gauge) coordinates of an image u

Miscellaneous

$\delta_{n,m}$	Kronecker delta, $\delta_{n,m} \coloneqq 1$ if $n = m$ and 0 else
x^T	transpose of the vector x

e_n	standard basis vector, $(e_n)_m \coloneqq \delta_{n,m}$
$\mathbb{1}$	vector of ones
$\otimes, \oslash$	pointwise multiplication and division
$\mathcal{O}$	$f \in \mathcal{O}(g)$ when $x \to 0$ iff $\exists\, c > 0$ so that for small x there is $f(x) \leq cg(x)$
$\langle x, y \rangle$	Euclidean scalar product, i.e. $\langle x, y \rangle \coloneqq \sum_n x_n y_n$
$\lvert x \rvert$	Euclidean norm, i.e. $\lvert x \rvert^2 \coloneqq \langle x, x \rangle = \sum_n x_n^2$
$\lvert x \rvert_\beta$	"smoothed" norm, i.e. for $\beta > 0$ $\lvert x \rvert_\beta^2 \coloneqq \lvert x \rvert^2 + \beta^2$
$\mathbb{B}_R(x), \mathbb{B}_R$	Ball of radius R around x, $\mathbb{B}_R \coloneqq \mathbb{B}_R(0)$
$p(u), p(f\lvert u)$	probability (density function) of/at f and the conditional probability of f given u
$\mathbb{E}(\mathtt{F})$	expectation of $\mathtt{F}$

Bibliography

[1] L. Alvarez, P.-L. Lions, and J.-M. Morel. "Image Selective Smoothing and Edge Detection by Nonlinear Diffusion II". *SIAM Journal on Numerical Analysis* 29.3 (1992), pp. 845–866.

[2] P. Arbelaez, C. Fowlkes, and D. Martin. *The Berkeley Segmentation Dataset [online]*. 2013. URL: `http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/` (visited on 01/03/2013).

[3] S. R. Arridge and A. Simmons. "Multi-Spectral Probabilistic Diffusion Using Bayesian Classification". *Scale-Space Theories in Computer Vision*. Ed. by B. M. ter Haar Romeny, L. Florack, J. J. Koenderink, and M. A. Viergever. Berlin: Springer, 1997, pp. 224–235.

[4] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*. Applied Mathematical Sciences. Springer, 2001.

[5] K. Baete, J. Nuyts, W. Van Paesschen, P. Suetens, and P. Dupont. "Anatomical-based FDG-PET Reconstruction for the Detection of Hypo-Metabolic Regions in Epilepsy". *IEEE Transactions on Medical Imaging* 23.4 (2004), pp. 510–9.

[6] D. L. Bailey, D. W. Townsend, P. E. Valk, and M. N. Maisey. *Positron Emission Tomography - Basic Sciences*. London: Springer, 2005.

[7] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé. "A Variational Model for P+XS Image Fusion". *International Journal of Computer Vision* 69.1 (2006), pp. 43–58.

[8] D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.

[9] H. H. Barrett and K. J. Myers. *Foundations of Image Science*. Wiley-Interscience, 2004, p. 1540.

[10] A. Bartels, P. Dülk, D. Trede, T. Alexandrov, and P. Maaß. "Compressed Sensing in Imaging Mass Spectrometry". *Inverse Problems* 29.12 (2013), p. 125015.

[11] S. Becker. *L-BFGS-B: Mex Wrapper [online]*. 2014. URL: `http://www.mathworks.co.uk/matlabcentral/fileexchange/35104-lbfgsb--l-bfgs-b--mex-wrapper` (visited on 02/11/2014).

[12] M. Bertero, H. Lantéri, and L. Zanni. "Iterative Image Reconstruction: A Point of View". *Proceedings of the Interdisciplinary Workshop on Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT)*. Pisa, 2007, pp. 1–25.

[13] B. Bilgic, V. K. Goyal, and E. Adalsteinsson. "Multi-Contrast Reconstruction with Bayesian Compressed Sensing". *Magnetic Resonance in Medicine* 66.6 (2011), pp. 1601–15.

[14] C. M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, 2006.

[15] P. Blomgren and T. F. Chan. "Color TV: Total Variation Methods for Restoration of Vector-Valued Images." *IEEE Transactions on Image Processing* 7.3 (1998), pp. 304–309.

[16] A. Bousse, S. Pedemonte, B. A. Thomas, K. Erlandsson, S. Ourselin, S. R. Arridge, and B. F. Hutton. "Markov Random Field and Gaussian Mixture for Segmented MRI-based Partial Volume Correction in PET". *Physics in Medicine and Biology* 57.20 (2012), pp. 6681–705.

[17] A. Bousse, S. Pedemonte, D. Kazantsev, S. Ourselin, S. R. Arridge, and B. F. Hutton. "Weighted MRI-based Bowsher Priors for SPECT Brain Image Reconstruction". *IEEE Nuclear Science Symposium and Medical Imaging Conference*. 2010, pp. 3519–3522.

[18] J. E. Bowsher, V. E. Johnson, T. G. Turkington, R. J. Jaszczak, C. E. Floyd, and R. E. Coleman. "Bayesian Reconstruction and Use of Anatomical A Priori Information for Emission Tomography". *IEEE Transactions on Medical Imaging* 15.5 (1996), pp. 673–686.

[19] J. E. Bowsher, H. Yuan, L. W. Hedlund, T. G. Turkington, G. Akabani, A. Badea, W. C. Kurylo, C. T. Wheeler, G. P. Cofer, M. W. Dewhirst, and G. A. Johnson. "Utilizing MRI Information to Estimate F18-FDG Distributions in Rat Flank Tumors". *IEEE Nuclear Science Symposium and Medical Imaging Conference*. 2004, pp. 2488–2492.

[20] K. Bredies and D. A. Lorenz. *Mathematische Bildverarbeitung: Einführung in Grundlagen und moderne Theorie (German)*. Vieweg+Teubner Verlag, 2011.

[21] K. Bredies, Y. Dong, and M. Hintermüller. "Spatially Dependent Regularization Parameter Selection in Total Generalized Variation Models for Image Restoration". *International Journal of Computer Mathematics* April 2015 (2012), pp. 1–15.

[22] X. Bresson and T. F. Chan. "Fast Dual Minimization of the Vectorial Total Variation Norm and Applications to Color Image Processing". *Inverse Problems and Imaging* 2.4 (2008), pp. 455–484.

[23] C. Brune. *4D Imaging in Tomography and Optical Nanoscopy*. PhD thesis. Westfaelische Wilhelms-Universitaet Muenster, 2010.

[24] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. "A Limited Memory Algorithm for Bound Constrained Optimization". *SIAM Journal of Scientific Computing* 16.5 (1995), pp. 1190–1208.

[25] R. H. Byrd, P. Lu, J. L. Morales, J. Nocedal, and C. Zhu. *L-BFGS-B: Software for Large-Scale Bound Constrained Optimization [online]*. 2014. URL: `http://www.ece.northwestern.edu/~nocedal/lbfgsb.html` (visited on 02/11/2014).

[26] C. L. Byrne. *Applied Iterative Methods*. A K Peters/CRC Press, 2007.

[27] E. J. Candès and J. K. Romberg. *l1-MAGIC: Recovery of Sparse Signals via Convex Programming*. Tech. rep. Caltech, 2005, pp. 1–19.

[28] E. J. Candès and J. K. Romberg. "Sparsity and Incoherence in Compressive Sampling". *Inverse Problems* 23.3 (2007), pp. 969–985.

[29] E. J. Candès, J. K. Romberg, and T. Tao. "Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information". *IEEE Transactions on Information Theory* 52.2 (2006), pp. 489–509.

[30] E. J. Candès, J. K. Romberg, and T. Tao. "Stable Signal Recovery from Incomplete and Inaccurate Measurements". *Communications on Pure and Applied Mathematics* LIX (2006), pp. 1207–1223.

[31] V. Caselles, B. Coll, and J.-M. Morel. "Geometry and Color in Natural Images". *Journal of Mathematical Imaging and Vision* 16.Section 2 (2002), pp. 89–105.

[32] C. Catana, A. R. Guimaraes, and B. R. Rosen. "PET and MR Imaging: The Odd Couple or a Match Made in Heaven?" *Journal of Nuclear Medicine* 54.5 (2013), pp. 815–24.

[33] C. Chan, R. Fulton, D. D. Feng, W. Cai, and S. Meikle. "An Anatomically based Regionally Adaptive Prior for MAP Reconstruction in Emission Tomography". *IEEE Nuclear Science Symposium and Medical Imaging Conference*. 2007, pp. 4137–4141.

[34] C. Chan, R. Fulton, D. D. Feng, and S. Meikle. "Regularized Image Reconstruction with an Anatomically Adaptive Prior for Positron Emission Tomography". *Physics in Medicine and Biology* 54.24 (2009), pp. 7379–400.

[35] T.-C. Chang, L. He, and T. Fang. "MR Image Reconstruction from Sparse Radial Samples Using Bregman Iteration". *International Society for Magnetic Resonance in Medicine*. Vol. 4. 2006, p. 696.

[36] Y. Chen, R. Ranftl, and T. Pock. "A Bi-Level View of Inpainting-based Image Compression". *19th Computer Vision Winter Workshop*. 2014.

[37] J. Cheng-Liao and J. Qi. "PET Image Reconstruction with Anatomical Edge Guided Level Set Prior". *Physics in Medicine and Biology* 56 (2011), pp. 6899–6918.

[38] J. Cheng-Liao and J. Qi. "Segmentation of Mouse Dynamic PET Images using a Multiphase Level Set Method". *Physics in Medicine and Biology* 55.21 (2010), pp. 6549–69.

[39] S. R. Cherry. "Multimodality in Vivo Imaging Systems: Twice the Power or Double the Trouble?" *Annual Review of Biomedical Engineering* 8 (2006), pp. 35–62.

[40] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, G. B. Pike, and A. C. Evans. "BrainWeb: Online Interface to a 3D MRI Simulated Brain Database". *NeuroImage* 5 (1997), p. 425.

[41] C. Comtat, P. E. Kinahan, J. A. Fessler, T. Beyer, D. W. Townsend, M. Defrise, and C. J. Michel. "Clinically Feasible Reconstruction of 3D Whole-Body PET/CT Data using Blurred Anatomical Labels". *Physics in Medicine and Biology* 47.1 (2002), pp. 1–20.

[42] D. Cremers, M. Rousson, and R. Deriche. “A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape”. *International Journal of Computer Vision* 72.2 (2006), pp. 195–215.

[43] S. Di Zenzo. “A Note on the Gradient of a Multi-Image”. *Computer Vision, Graphics and Image Processing* 33 (1986), pp. 116–125.

[44] O. Dietrich, J. G. Raya, S. B. Reeder, M. F. Reiser, and S. O. Schoenberg. “Measurement of Signal-to-Noise Ratios in MR Images: Influence of Multichannel Coils, Parallel Imaging, and Reconstruction filters”. *Journal of Magnetic Resonance Imaging* 26.2 (2007), pp. 375–85.

[45] Y. Dong, M. Hintermüller, and M. M. Rincon-Camacho. “Automated Regularization Parameter Selection in Multi-Scale Total Variation Models for Image Restoration”. *Journal of Mathematical Imaging and Vision* 40.1 (2011), pp. 82–104.

[46] D. L. Donoho. “Compressed Sensing”. *IEEE Transactions on Information Theory* 52.4 (2006), pp. 1289–1306.

[47] J. J. Duistermaat and J. A. C. Kolk. *Multidimensional Real Analysis I: Differentiation*. New York: Cambridge University Press, 2004.

[48] J. J. Duistermaat and J. A. C. Kolk. *Multidimensional Real Analysis II: Integration*. New York: Cambridge University Press, 2004.

[49] M. J. Ehrhardt and S. R. Arridge. *Software for Parallel Level Sets [online]*. 2013. URL: `http://www0.cs.ucl.ac.uk/staff/ehrhardt/software.html`.

[50] M. J. Ehrhardt and S. R. Arridge. “Vector-Valued Image Processing by Parallel Level Sets”. *IEEE Transactions on Image Processing* 23.1 (2014), pp. 9–18.

[51] M. J. Ehrhardt, K. Thielemans, L. Pizarro, D. Atkinson, S. Ourselin, B. F. Hutton, and S. R. Arridge. “Joint Reconstruction of PET-MRI by exploiting Structural Similarity”. *Inverse Problems* 31 (2015), p. 015001.

[52] M. J. Ehrhardt, K. Thielemans, L. Pizarro, P. Markiewicz, D. Atkinson, S. Ourselin, B. F. Hutton, and S. R. Arridge. “Joint Reconstruction of PET-MRI by Parallel Level Sets”. *IEEE Nuclear Science Symposium and Medical Imaging Conference*. 2014.

[53] Y. C. Eldar and G. Kutyniok. *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[54] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Springer, 1996.

[55] C. L. Epstein. *Introduction to the Mathematics of Medical Imaging*. Philadelphia: Society for Industrial and Applied Mathematics, 2008.

[56] K. Erlandsson, I. Buvat, P. H. Pretorius, B. A. Thomas, and B. F. Hutton. "A Review of Partial Volume Correction Techniques for Emission Tomography and their Applications in Neurology, Cardiology and Oncology". *Physics in Medicine and Biology* 57.21 (2012), R119–59.

[57] V. Estellers, J. Thiran, and X. Bresson. "Enhanced Compressed Sensing Recovery With Level Set Normals". *IEEE Transactions on Image Processing* 22.7 (2013), pp. 2611–2626.

[58] F. J. Estrada. *Image Denoising Benchmark [online]*. 2013. URL: `http://www.cs.utoronto.ca/~strider/Denoise/Benchmark/index.html` (visited on 01/03/2013).

[59] F. Fang, F. Li, C. Shen, and G. Zhang. "A Variational Approach for Pan-Sharpening". *IEEE Transactions on Image Processing* 22.7 (2013), pp. 2822–2834.

[60] H. O. Fattorini. *Infinite Dimensional Optimization and Control Theory*. Cambridge University Press, 1999.

[61] J. A. Fessler and D. C. Noll. "Iterative Image Reconstruction in MRI with Separate Magnitude and Phase Regularization". *International Symposium on Biomedical Imaging*. 2004, pp. 209–212.

[62] J. A. Fessler, I. Elbakri, P. Sukovic, and N. H. Clinthorne. "Maximum-likelihood dual-energy tomographic image reconstruction". *SPIE: Medical Imaging*. Vol. 4684. 2002, pp. 1–25.

[63] L. A. Gallardo and M. A. Meju. "Characterization of Heterogeneous Near-Surface Materials by Joint 2D Inversion of DC Resistivity and Seismic Data". *Geophysical Research Letters* 30.13 (2003), p. 1658.

[64] L. A. Gallardo and M. A. Meju. "Joint Two-Dimensional DC Resistivity and Seismic Travel Time Inversion with Cross-Gradients Constraints". *Journal of Geophysical Research* 109.B3 (2004), pp. 1–11.

[65] L. A. Gallardo and M. A. Meju. “Structure-Coupled Multiphysics Imaging in Geophysical Sciences”. *Reviews of Geophysics* 49 (2011), pp. 1–19.

[66] U. Gamper, P. Boesiger, and S. Kozerke. “Compressed Sensing in Dynamic MRI”. *Magnetic Resonance in Medicine* 59.2 (2008), pp. 365–73.

[67] B. Goldluecke, E. Strekalovskiy, and D. Cremers. “The Natural Vectorial Total Variation Which Arises from Geometric Measure Theory”. *SIAM Journal on Imaging Sciences* 5.2 (2012), pp. 537–563.

[68] P. J. Green. “Bayesian Reconstructions from Emission Tomography Data using a Modified EM Algorithm”. *IEEE Transactions on Medical Imaging* 9.893 (1990).

[69] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. “Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA)”. *Magnetic Resonance in Medicine* 47 (2002), pp. 1202–1210.

[70] H. Gudbjartsson and S. Patz. “The Rician Distribution of Noisy MRI Data”. *Magnetic Resonance in Medicine* 43.6 (1995), pp. 910–914.

[71] M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, and M. Unser. “Realistic Analytical Phantoms for Parallel Magnetic Resonance Imaging”. *IEEE Transactions on Medical Imaging* 31.3 (2012), pp. 626–636.

[72] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau. “Color Plane Interpolation using Alternating Projections”. *IEEE Transactions on Image Processing* 11.9 (2002), pp. 997–1013.

[73] B. K. Gunturk, J. Glotzbach, Y. Altunbasak, R. W. Schafer, and R. M. Mersereau. “Demosaicking: Color Filter Array Interpolation”. *IEEE Signal Processing Magazine* (2005), pp. 44–54.

[74] E. Haber and M. Holtzman-Gazit. “Model Fusion and Joint Inversion”. *Surveys in Geophysics* 34 (2013), pp. 675–695.

[75] E. Haber and J. Modersitzki. “Intensity Gradient based Registration and Fusion of Multi-Modal Images”. *Medical Image Computing and Computer-Assisted Intervention*. Vol. 46. 3. Berlin Heidelberg: Springer-Verlag, 2006, pp. 726–733.

[76] E. Haber and D. W. Oldenburg. “Joint Inversion: A Structural Approach”. *Inverse Problems* 13 (1997), pp. 63–77.

[77] E. Haber, K. van den Doel, and L. Horesh. "Optimal Design of Simultaneous Source Encoding". *Inverse Problems in Science and Engineering* 23.5 (2014), pp. 780–797.

[78] M. Hintermüller and M. M. Rincon-Camacho. "Expected Absolute Value Estimators for a Spatially Adapted Regularization Parameter Choice Rule in L1-TV-based Image Restoration". *Inverse Problems* 26.8 (2010), p. 085005.

[79] W. S. Hoge, D. H. Brooks, B. Madore, and W. E. Kyriakos. "A Tour of Accelerated Parallel MR Imaging from a Linear Systems Perspective". *Concepts in Magnetic Resonance Part A* 27A.1 (2005), pp. 17–37.

[80] K. M. Holt. "Total Nuclear Variation and Jacobian Extensions of Total Variation for Vector Fields". *IEEE Transactions on Image Processing* 23.9 (2014), pp. 3975–3989.

[81] J. Hunter and B. Nachtergaele. *Applied Analysis*. World Scientific Publishing Company Incorporated, 2001.

[82] B. F. Hutton, B. A. Thomas, K. Erlandsson, A. Bousse, A Reilhac-Laborde, D. Kazantsev, S. Pedemonte, K. Vunckx, S. R. Arridge, and S. Ourselin. "What Approach to Brain Partial Volume Correction is best for PET/MRI?" *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 702 (2013), pp. 29–33.

[83] B. Jin, D. A. Lorenz, and S. Schiffler. "Elastic-Net Regularization: Error Estimates and Active Set Methods". *Inverse Problems* 25.11 (2009), p. 115022.

[84] M. S. Judenhofer, H. F. Wehrl, D. F. Newport, C. Catana, S. B. Siegel, M. Becker, A. Thielscher, M. Kneilling, M. P. Lichy, M. Eichner, K. Klingel, G. Reischl, S. Widmaier, M. Röcken, R. E. Nutt, H.-J. Machulla, K. Uludag, S. R. Cherry, C. D. Claussen, and B. J. Pichler. "Simultaneous PET-MRI: A New Approach for Functional and Morphological Imaging". *Nature Medicine* 14.4 (2008), pp. 459–65.

[85] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Vol. 160. Springer, 2006.

[86] J. P. Kaipio, V. Kolehmainen, M. Vauhkonen, and E. Somersalo. "Inverse Problems with Structural Prior Information". *Inverse Problems* 15.3 (1999), pp. 713–729.

[87] D. Kazantsev, S. R. Arridge, S. Pedemonte, A. Bousse, K. Erlandsson, B. F. Hutton, and S. Ourselin. "An Anatomically Driven Anisotropic Diffusion Filtering Method for 3D SPECT Reconstruction". *Physics in Medicine and Biology* 57.12 (2012), pp. 3793–3810.

[88] D. Kazantsev, A. Bousse, S. Pedemonte, S. R. Arridge, and B. F. Hutton. "Edge Preserving Bowsher Prior with Nonlocal Weighting for 3D SPECT Reconstruction". *IEEE Nuclear Science Symposium and Medical Imaging Conference*. 2011, pp. 1158–1161.

[89] D. Kazantsev, W. R. B. Lionheart, P. J. Withers, and P. D. Lee. "Multimodal Image Reconstruction using Supplementary Structural Information in Total Variation Regularization". *Sensing and Imaging* 15.1 (2014), p. 97.

[90] H. Kekkonen, M. Lassas, and S. Siltanen. "Analysis of Regularized Inversion of Data Corrupted by White Gaussian Noise". *Inverse Problems* 30 (2014), p. 045009.

[91] R. Kimmel, R. Malladi, and N. Sochen. "Images as Embedded Maps and Minimal Surfaces: Movies, Color, Texture, and Volumetric Medical Images". *International Journal of Computer Vision* 39.2 (2000), pp. 111–129.

[92] F. Knoll, T. Koesters, R. Otazo, K. T. Block, L. Feng, K. Vunckx, D. Faul, J. Nuyts, F. Boada, and D. K. Sodickson. "Joint Reconstruction of Simultaneously acquired MR-PET Data with Multi Sensor Compressed Sensing based on a Joint Sparsity Constraint". *EJNMMI Physics* 1.Suppl 1 (2014), A26.

[93] F. Knoll, C. Clason, K. Bredies, M. Uecker, and R. Stollberger. "Parallel Imaging with Nonlinear Reconstruction using Variational Penalties". *Magnetic Resonance in Medicine* 67.1 (2012), pp. 34–41.

[94] F. Knoll, T. Koesters, R. Otazo, F. Boada, and D. K. Sodickson. "Simultaneous MR-PET Reconstruction using Multi Sensor Compressed Sensing and Joint Sparsity". *International Society for Magnetic Resonance in Medicine*. Vol. 22. 82. 2014.

[95] F. Knoll, Y. Dong, C. Langskammer, M. Hintermüller, and R. Stollberger. "Total Variation Denoising with Spatially Dependent Regularization". *Proc. Intl. Soc. Mag. Reson. Med.* Vol. 18. 2010, p. 5088.

[96] D. J. Larkman and R. G. Nunes. “Parallel Magnetic Resonance Imaging”. *Physics in Medicine and Biology* 52.7 (2007), R15–55.

[97] R. M. Leahy and X. Yan. “Incorporation of Anatomical MR Data for Improved Functional Imaging with PET”. *Information Processing in Medical Imaging*. Springer. 1991, pp. 105–120.

[98] M. S. Lehtinen, L. Paivarinta, and E. Somersalo. “Linear Inverse Problems for Generalised Random Variables”. *Inverse Problems* 5 (1989), pp. 599–612.

[99] Z.-P. Liang and P. C. Lauterbur. *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*. IEEE Press Series in Biomedical Engineering. SPIE Optical Engineering Press, 2000.

[100] J. Liao and J. Qi. “PET Image Reconstruction with Anatomical Prior Using Multiphase Level Set Method”. *IEEE Nuclear Science Symposium and Medical Imaging Conference*. 2007, pp. 4163–4168.

[101] B. Liu, L. Ying, M. Steckner, J. Xie, and J. Sheng. “Regularized SENSE Reconstruction using Iteratively Refined Total Variation Method”. *IEEE International Symposium on Biomedical Imaging*. 2007, pp. 121–124.

[102] S. Lu and S. V. Pereverzev. “Multi-Parameter Regularization and its Numerical Realization”. *Numerische Mathematik* 118.1 (2011), pp. 1–31.

[103] M. Lustig. *Sparse MRI*. PhD. Stanford University, 2008, pp. 1–159.

[104] M. Lustig, D. L. Donoho, and J. M. Pauly. “Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging”. *Magnetic Resonance in Medicine* 58.6 (2007), pp. 1182–1195.

[105] A. Macovski. “Noise in MRI”. *Magnetic Resonance in Medicine* 36.3 (1996), pp. 494–497.

[106] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. “Multimodality Image Registration by Maximization of Mutual Information”. *IEEE Transactions on Medical Imaging* 16.2 (1997), pp. 187–98.

[107] D. Martin, C. Fowlkes, D. Tal, and J. Malik. “A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics”. *Proc. 8th Int’l Conf. Computer Vision*. Vol. 2. 2001, pp. 416–423.

[108] D. W. McRobbie, E. A. Moore, M. J. Graves, and M. R. Prince. *MRI - From Picture to Proton*. Cambridge University Press, 2006.

[109] S. K. Mishra and G. Giorgi. *Invexity and Optimization*. Springer, 2008.

[110] M. Möller, T. Wittman, A. L. Bertozzi, and M. Burger. "A Variational Approach for Sharpening High Dimensional Images". *SIAM Journal on Imaging Sciences* 5.1 (2012), pp. 150–178.

[111] M. Möller, T. Wittman, and A. L. Bertozzi. "A Variational Approach to Hyperspectral Image Fusion". *SPIE: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV*. Ed. by S. S. Shen and P. E. Lewis. Vol. 7334. 2009, 73341E–73341E–10.

[112] M. Möller, E.-M. Brinkmann, M. Burger, and T. Seybold. "Color Bregman TV". *SIAM Journal on Imaging Sciences* 7.4 (2014), pp. 2771–2806.

[113] J. L. Morales and J. Nocedal. "Remark on "Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound Constrained Optimization"". *ACM Transactions on Mathematical Software* 38.1 (2011), pp. 1–4.

[114] F. Natterer. *The Mathematics of Computerized Tomography*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 2001.

[115] F. Natterer and F. Wübbeling. *Mathematical Methods in Image Reconstruction*. SIAM Monographs on Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics, 2001.

[116] C. Niculescu and L.-E. Persson. *Convex Functions and their Applications: A Contemporary Approach*. Springer, 2005.

[117] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 2000.

[118] J. Nuyts. "The Use of Mutual Information and Joint Entropy for Anatomical Priors in Emission Tomography". *IEEE Nuclear Science Symposium and Medical Imaging Conference*. IEEE, 2007, pp. 4149–4154.

[119] J. M. Ollinger and J. A. Fessler. "Positron-Emisson Tomography". *IEEE Signal Processing Magazine* (1997).

[120] S. Osher and L. I. Rudin. "Feature-Oriented Image Enhancement using Shock Filters". *SIAM Journal on Numerical Analysis* 27.4 (1990), pp. 919–940.

[121] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. "An Iterative Regularization Method for Total Variation-based Image Restoration". *SIAM Multicale Modelling and Simulation* 4.2 (2005), pp. 460–489.

[122] S. Pedemonte and S. Ourselin. *NiftyRec 2.0 Tomography Toolbox [online]*. 2015. URL: `http://niftyrec.scienceontheweb.net/wordpress/` (visited on 06/12/2013).

[123] S. Pedemonte, A. Bousse, B. F. Hutton, S. R. Arridge, and S. Ourselin. "4-D Generative Model for PET/MRI Reconstruction". *Medical Image Computing and Computer-Assisted Intervention*. Springer, 2011, pp. 581–588.

[124] S. Pedemonte, A. Bousse, K. Erlandsson, M. Modat, S. R. Arridge, B. F. Hutton, and S. Ourselin. "GPU Accelerated Rotation-based Emission Tomography Reconstruction". *IEEE Nuclear Science Symposium and Medical Imaging Conference*. 2010, pp. 2657–2661.

[125] S. Pedemonte, A. Bousse, B. F. Hutton, S. R. Arridge, and S. Ourselin. "Probabilistic Graphical Model of SPECT/MRI". *Machine Learning in Medical Imaging*. 2011, pp. 167–174.

[126] S. Pedemonte, M. J. Cardoso, S. R. Arridge, B. F. Hutton, and S. Ourselin. "Steady-State Model of the Radio-Pharmaceutical Uptake for MR-PET". *Medical Image Computing and Computer-Assisted Intervention*. 2012, pp. 289–297.

[127] P. Perona and J. Malik. "Scale-Space and Edge Detection using Anisotropic Diffusion". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12.7 (1990), pp. 629–639.

[128] Y. Petibon, G. El Fakhri, R. Nezafat, N. Johnson, T. Brady, and J. Ouyang. "Towards Coronary Plaque Imaging using Simultaneous PET-MR: A Simulation Study". *Physics in Medicine and Biology* 59.5 (2014), pp. 1203–22.

[129] B. J. Pichler, H. F. Wehrl, A. Kolb, and M. S. Judenhofer. "Positron Emission Tomography/Magnetic Resonance Imaging: The Next Generation of Multimodality Imaging?" *Seminars in Nuclear Medicine* 38.3 (2008), pp. 199–208.

[130] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. "SENSE: Sensitivity Encoding for Fast MRI". *Magnetic Resonance in Medicine* 42.5 (1999), pp. 952–62.

[131] A. Rangarajan, I.-T. Hsiao, and G. Gindi. “A Bayesian Joint Mixture Framework for the Integration of Anatomical Information in Functional Image Reconstruction”. *Journal of Mathematical Imaging and Vision* 12.3 (2000), pp. 199–217.

[132] S. Ravishankar and Y. Bresler. “MR Image Reconstruction From Highly Undersampled K-Space Data by Dictionary Learning”. *IEEE Transactions on Medical Imaging* 30.5 (2011), pp. 1028–41.

[133] D. Rigie and P. La Riviere. “Joint Reconstruction of Multi-Channel, Spectral CT Data via Constrained Total Nuclear Variation Minimization”. *Physics in Medicine and Biology* 60 (2015), pp. 1741–1762.

[134] L. I. Rudin, S. Osher, and E. Fatemi. “Nonlinear Total Variation based Noise Removal Algorithms”. *Physica D: Nonlinear Phenomena* 60.1 (1992), pp. 259–268.

[135] G. Sapiro and D. L. Ringach. “Anisotropic Diffusion of Multivalued Images with Applications to Color Filtering”. *IEEE Transactions on Image Processing* 5.11 (1996), pp. 1582–6.

[136] A. Sawatzky, C. Brune, T. Koesters, F. Wübbeling, and M. Burger. “Level Set and PDE based Reconstruction Methods in Imaging”. *Level Set and PDE Based Reconstruction Methods in Imaging*. Vol. 2090. Lecture Notes in Mathematics. Springer International Publishing Switzerland, 2013.

[137] L. A. Shepp and Y Vardi. “Maximum Likelihood Reconstruction for Emission Tomography”. *IEEE Transactions on Medical Imaging* 1.2 (1982), pp. 113–22.

[138] K. Shmueli, J. A. de Zwart, P. van Gelderen, T.-Q. Li, S. J. Dodd, and J. H. Duyn. “Magnetic Susceptibility Mapping of Brain Tissue in Vivo using MRI Phase Data”. *Magnetic Resonance in Medicine* 62.6 (2009), pp. 1510–22.

[139] N. Sochen, R. Kimmel, and R. Malladi. “A General Framework for Low Level Vision”. *IEEE Transactions on Image Processing* 7.3 (1998), pp. 310–318.

[140] S. Somayajula, E. Asma, and R. M. Leahy. “PET Image Reconstruction using Anatomical Information through Mutual Information based Priors”. *IEEE Nuclear Science Symposium and Medical Imaging Conference*. 2005, pp. 2722–2726.

[141] S. Somayajula, C. Panagiotou, A. Rangarajan, Q. Li, S. R. Arridge, and R. M. Leahy. “PET Image Reconstruction using Information Theoretic Anatomical Priors”. *IEEE Transactions on Medical Imaging* 30.3 (2011), pp. 537–549.

[142] J. Tang and A. Rahmim. "Bayesian PET Image Reconstruction Incorporating Anato-Functional Joint Entropy". *Physics in Medicine and Biology* 54.23 (2009), pp. 7063–75.

[143] K. Thielemans, C. Tsoumpas, S. Mustafovic, T. Beisel, P. Aguiar, N. Dikaios, and M. W. Jacobson. *STIR: Software for Tomographic Image Reconstruction [online]*. 2015. URL: `http://stir.sourceforge.net` (visited on 10/03/2014).

[144] K. Thielemans, C. Tsoumpas, S. Mustafovic, T. Beisel, P. Aguiar, N. Dikaios, and M. W. Jacobson. "STIR: Software for Tomographic Image Reconstruction Release 2". *Physics in Medicine and Biology* 57 (2012), pp. 867–883.

[145] D. W. Townsend. "Multimodality Imaging of Structure and Function". *Physics in Medicine and Biology* 53.4 (2008), R1–R39.

[146] D. Tschumperlé and R. Deriche. "Vector-Valued Image Regularization with PDEs: A Common Framework for Different Applications." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.4 (2005), pp. 506–17.

[147] G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. "Robust and Efficient Parametric Face Alignment". *IEEE International Conference on Computer Vision*. 2011, pp. 1847–1854.

[148] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig. "ESPIRiT - An Eigenvalue Approach to Autocalibrating Parallel MRI: Where SENSE meets GRAPPA". *Magnetic Resonance in Medicine* 71 (2014), pp. 990–1001.

[149] M. Uecker, T. Hohage, K. T. Block, and J. Frahm. "Image Reconstruction by Regularized Nonlinear Inversion - Joint Estimation of Coil Sensitivities and Image Content". *Magnetic Resonance in Medicine* 60 (2008), pp. 674–682.

[150] T. Valkonen. "A Primal-Dual Hybrid Gradient Method for Nonlinear Operators with Applications to MRI". *Inverse Problems* 30.5 (2014), p. 055012.

[151] S. S. Vasanawala, M. J. Murphy, M. T. Alley, K. Keutzer, P. Lai, J. M. Pauly, and M. Lustig. "Practical Parallel Imaging Compressed Sensing MRI: Summary of Two Years of Experience in Accelerating Body MRI of Pediatric Patients". *International Symposium on Biomedical Imaging*. 2011, pp. 1039–1043.

[152] K. Vunckx and J. Nuyts. “Heuristic Modification of an Anatomical Markov Prior Improves its Performance”. *IEEE Nuclear Science Symposium and Medical Imaging Conference*. 2010, pp. 3262–3266.

[153] K. Vunckx, A. Atre, K. Baete, A. Reilhac, C. M. Deroose, K. Van Laere, and J. Nuyts. “Evaluation of Three MRI-based Anatomical Priors for Quantitative PET Brain Imaging”. *IEEE Transactions on Medical Imaging* 31.3 (2012), pp. 599–612.

[154] G. Wang, J. Zhang, H. Gao, V. Weir, H. Yu, W. Cong, X. Xu, H. Shen, J. Bennett, M. Furth, Y. Wang, and M. Vannier. “Towards Omni-Tomography–Grand Fusion of Multiple Modalities for Simultaneous Interior Tomography”. *PloS one* 7.6 (2012), e39700.

[155] W. Wang, S. Lu, H. Mao, and J. Cheng. “Multi-Parameter Tikhonov Regularization with the l0 Sparsity Constraint”. *Inverse Problems* 29.6 (2013), p. 065018.

[156] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. “Image Quality Assessment: From Error Visibility to Structural Similarity”. *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–12.

[157] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. *The SSIM Index for Image Quality Assessment [online]*. 2013. URL: `https://ece.uwaterloo.ca/~z70wang/research/ssim/` (visited on 01/03/2013).

[158] J. Weickert. *Anisotropic Diffusion in Image Processing*. Stuttgart: Teubner, 1998.

[159] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. “Multi-Modal Volume Registration by Maximization of Mutual Information”. *Medical Image Analysis* 1.1 (1996), pp. 35–51.

[160] R. T. Whitaker. “Geometry-Limited Diffusion in the Characterization of Geometric Patches in Images”. *CVGIP: Image Understanding* 57.1 (1993), pp. 111–120.

[161] H. Zaidi and A. Del Guerra. “An Outlook on Future Design of Hybrid PET/MRI Systems”. *Medical Physics* 38.10 (2011), pp. 5667–89.

[162] E. Zeidler. *Applied Functional Analysis: Main Principles and Their Applications*. Applied Mathematical Sciences Series. Springer-Verlag, 1995.

[163] E. Zeidler. *Nonlinear Functional Analysis and Its Applications I: Fixed Point Theorems*. Springer-Verlag, 1985.

[164] F. Zhao, D. C. Noll, J.-F. Nielsen, and J. A. Fessler. "Separate Magnitude and Phase Regularization via Compressed Sensing". *IEEE Transactions on Medical Imaging* 31.9 (2012), pp. 1713–23.

[165] M. V. W. Zibetti and A. R. De Pierro. "Separate Magnitude and Phase Regularization in MRI with Incomplete Data: Preliminary Results". *International Symposium on Biomedical Imaging*. 2010, pp. 736–739.